

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
14 February 2002 (14.02.2002)

PCT

(10) International Publication Number
WO 02/12562 A2

(51) International Patent Classification⁷: **C12Q 1/68**,
C12N 5/10, G06F 17/00, C12N 15/12, 9/20, C07K 16/40

(21) International Application Number: PCT/US01/24663

(22) International Filing Date: 6 August 2001 (06.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/223,179 4 August 2000 (04.08.2000) US

(71) Applicant (for all designated States except US): **GENAIS-
SANCE PHARMACEUTICALS, INC.** [US/US]; A Cor-
poration of the State of Delaware, Five Science Park, New
Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **KAZEMI, Amir**
[US/US]; 53 Southwind Drive, Wallingford, CT US (US).
KLIEM, Stefanie E. [DE/DE]; Kiefernweg 37, 61440
Oberusel (DE). **KOSHY, Beena** [IN/US]; 1298 Hartford
Turnpike, Apt. 11B, North Haven, CT 06473 (US).

(74) Agents: **FIELD, Gisela M.** et al.; Genaisance Pharma-
ceuticals, Inc., Five Science Park, New Haven, CT 06511
(US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,
ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,
TG).

Published:

- without international search report and to be republished
upon receipt of that report
- with sequence listing part of description published sepa-
rately in electronic form and available upon request from
the International Bureau

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(54) Title: HAPLOTYPES OF THE PLA2G1B GENE

(57) Abstract: Novel genetic variants of the Phospholipase A2, Group IB (Pancreas) (PLA2G1B) gene are described. Various genotypes, haplotypes, and haplotype pairs that exist in the general United States population are disclosed for the PLA2G1B gene. Compositions and methods for haplotyping and/or genotyping the PLA2G1B gene in an individual are also disclosed. Polynucleotides defined by the haplotypes disclosed herein are also described.



WO 02/12562 A2

HAPLOTYPES OF THE PLA2G1B GENE

RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/223,179 filed August 4, 2000.

FIELD OF THE INVENTION

This invention relates to variation in genes that encode pharmaceutically-important proteins. In particular, this invention provides genetic variants of the human phospholipase A2, group IB (pancreas) (PLA2G1B) gene and methods for identifying which variant(s) of this gene is/are possessed by an individual.

BACKGROUND OF THE INVENTION

Current methods for identifying pharmaceuticals to treat disease often start by identifying, cloning, and expressing an important target protein related to the disease. A determination of whether an agonist or antagonist is needed to produce an effect that may benefit a patient with the disease is then made. Then, vast numbers of compounds are screened against the target protein to find new potential drugs. The desired outcome of this process is a lead compound that is specific for the target, thereby reducing the incidence of the undesired side effects usually caused by activity at non-intended targets. The lead compound identified in this screening process then undergoes further *in vitro* and *in vivo* testing to determine its absorption, disposition, metabolism and toxicological profiles. Typically, this testing involves use of cell lines and animal models with limited, if any, genetic diversity.

What this approach fails to consider, however, is that natural genetic variability exists between individuals in any and every population with respect to pharmaceutically-important proteins, including the protein targets of candidate drugs, the enzymes that metabolize these drugs and the proteins whose activity is modulated by such drug targets. Subtle alteration(s) in the primary nucleotide sequence of a gene encoding a pharmaceutically-important protein may be manifested as significant variation in expression, structure and/or function of the protein. Such alterations may explain the relatively high degree of uncertainty inherent in the treatment of individuals with a drug whose design is based upon a single representative example of the target or enzyme(s) involved in metabolizing the drug. For example, it is well-established that some drugs frequently have lower efficacy in some individuals than others, which means such individuals and their physicians must weigh the possible benefit of a larger dosage against a greater risk of side effects. Also, there is significant variation in how well people metabolize drugs and other exogenous chemicals, resulting in substantial interindividual variation in the toxicity and/or efficacy of such exogenous substances (Evans et al., 1999, *Science* 286:487-491). This variability in efficacy or toxicity of a drug in genetically-diverse patients makes many drugs ineffective or even dangerous in certain groups of the population, leading to the failure of such drugs in clinical trials or their early withdrawal from the market even though they could be highly beneficial for other groups in the population. This problem significantly increases the time and cost of drug

discovery and development, which is a matter of great public concern.

It is well-recognized by pharmaceutical scientists that considering the impact of the genetic variability of pharmaceutically-important proteins in the early phases of drug discovery and development is likely to reduce the failure rate of candidate and approved drugs (Marshall A 1997 *Nature Biotech* **15**:1249-52; Kleyn PW et al. 1998 *Science* **281**: 1820-21; Kola I 1999 *Curr Opin Biotech* **10**:589-92; Hill AVS et al. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 62-76; Meyer U.A. 1999 in *Evolution in Health and Disease* Stearns SS (Ed.) Oxford University Press, New York, pp 41-49; Kalow W et al. 1999 *Clin. Pharm. Therap.* **66**:445-7; Marshall, E 1999 *Science* **284**:406-7; Judson R et al. 2000 *Pharmacogenomics* **1**:1-12; Roses AD 2000 *Nature* **405**:857-65). However, in practice this has been difficult to do, in large part because of the time and cost required for discovering the amount of genetic variation that exists in the population (Chakravarti A 1998 *Nature Genet* **19**:216-7; Wang DG et al 1998 *Science* **280**:1077-82; Chakravarti A 1999 *Nat Genet* **21**:56-60 (suppl); Stephens JC 1999 *Mol. Diagnosis* **4**:309-317; Kwok PY and Gu S 1999 *Mol. Med. Today* **5**:538-43; Davidson S 2000 *Nature Biotech* **18**:1134-5).

The standard for measuring genetic variation among individuals is the haplotype, which is the ordered combination of polymorphisms in the sequence of each form of a gene that exists in the population. Because haplotypes represent the variation across each form of a gene, they provide a more accurate and reliable measurement of genetic variation than individual polymorphisms. For example, while specific variations in gene sequences have been associated with a particular phenotype such as disease susceptibility (Roses AD *supra*; Ulbrecht M et al. 2000 *Am J Respir Crit Care Med* **161**: 469-74) and drug response (Wolfe CR et al. 2000 *BMJ* **320**:987-90; Dahl BS 1997 *Acta Psychiatr Scand* **96** (Suppl 391): 14-21), in many other cases an individual polymorphism may be found in a variety of genomic backgrounds, i.e., different haplotypes, and therefore shows no definitive coupling between the polymorphism and the causative site for the phenotype (Clark AG et al. 1998 *Am J Hum Genet* **63**:595-612; Ulbrecht M et al. 2000 *supra*; Drysdale et al. 2000 *PNAS* **97**:10483-10488). Thus, there is an unmet need in the pharmaceutical industry for information on what haplotypes exist in the population for pharmaceutically-important genes. Such haplotype information would be useful in improving the efficiency and output of several steps in the drug discovery and development process, including target validation, identifying lead compounds, and early phase clinical trials (Marshall et al., *supra*).

One pharmaceutically-important gene for the treatment of pancreatitis and pancreatic cancer is the phospholipase A2, group IB (pancreas) (PLA2G1B) gene or its encoded product. Phospholipase A2 enzymes regulate the synthesis of arachidonic acid and their metabolites. There are three major groups of phospholipase A2 (PLA(2)): group I, also called pancreatic PLA(2) (PLA(2)-I), and herein referred to as PLA2G1B; group II, referred to as secretory non-pancreatic or synovial or platelet PLA(2) (PLA(2)-II); and group IV, referred to as cytosolic PLA(2) (PLA(2)-IV) (Kashiwagi et al., *Gut* 1999 Oct;**45**(4):605-12). PLA2G1B catalyzes the release of fatty acids from glycerol-3-phosphocholines. Cholesterol absorption from bile acid micelles is suppressed by phosphatidylcholine

(PC) in the micelles. PLA2G1B enhances cholesterol absorption from PC-containing micelles by eliminating micellar phosphatidylcholine through its lipolytic action. PLA2G1B inhibitors are therefore suggested to act as a novel class of cholesterol absorption inhibitors for therapeutic use (Homan and Hamelhele *J Lipid Res* 1998 39:1197-209).

Recent studies have shown that PLA2G1B stimulates the growth of human pancreatic cell lines by activating the mitogen-activated protein kinase pathway (Kinoshita et al., *FEBS Lett* 1997 407:343-6). Hongo et al. (*Immunopharmacol Immunotoxicol* 1999 21:717-26) investigated the importance of PLA2G1B for bacterial translocation which is considered to be one of the causes of acute pancreatitis. They found that when male rats were administered PLA2G1B intraperitoneally, the mesenteric lymph node (MLN) cell cultures of these rats had a high level of enteric bacteria. There were fewer positive MLN cultures in rats that were given nafamostat mesilate, a protease inhibitor, before the intraperitoneal injection of PLA2G1B. Hongo et al. (*supra*) suggested that PLA2G1B plays an important role in bacterial translocation in acute pancreatitis and the translocation can be blocked by administering protease inhibitors.

Recently, Friess et al. (*Ann Surg* 2001 233(2):204-12) et al. have shown that the different groups of phospholipase A2 are differentially expressed in individuals with acute pancreatitis. Relative to controls, PLA2G1B mRNA expression was 8.9-fold decreased in patients with acute pancreatitis. By contrast, phospholipase A2-II (7.8-fold) and phospholipase A2-IV (8.1-fold) mRNA levels were increased in individuals with this disorder. In rat models of acute pancreatitis, there was a similar increase in phospholipase A2-IV levels, suggesting that the phospholipase A2-II and A2-IV groups are involved in regulating the inflammatory response in this disease (Friess et al., *supra*).

In vitro and *in vivo* studies have shown that PLA2G1B may be associated with pancreatic cancer. PLA2G1B was demonstrated to be capable of stimulating growth of the human pancreatic cancer cell line MIAPaCa-2, while the pro-form of the enzyme did not stimulate growth of these cells (Kashiwagi et al., *supra*). Immunohistochemical analysis has shown that PLA2G1B is localized in over 80% of pancreatic ductal carcinomas examined. In the majority of these immunopositive samples, a granular cytoplasmic pattern was observed, which was not seen in normal cells (Kiyohara et al. *Int J Pancreatol* 1993 13(1):49-57).

The phospholipase A2, group IB (pancreas) gene is located on chromosome 12q23-q24.1 and contains 4 exons that encode a 148 amino acid protein. A reference sequence for the PLA2G1B gene is shown in the contiguous lines of Figure 1 (GenBank Reference No. 4762747; SEQ ID NO: 1). Reference sequences for the coding sequence (GenBank Accession No. NM_000928.1) and protein are shown in Figures 2 (SEQ ID NO: 2) and 3 (SEQ ID NO: 3), respectively.

Several single nucleotide polymorphisms in the PLA2G1B have been reported in the NCBI SNP database including a polymorphism of guanine or adenine at a position corresponding to nucleotide 3968 in Figure 1 (Halushka et al. *Nat. Genet.* 1999; 22:239-247).

Because of the potential for variation in the PLA2G1B gene to affect the expression and function of the encoded protein, it would be useful to know whether additional polymorphisms exist in

the PLA2G1B gene, as well as how such polymorphisms are combined in different copies of the gene. Such information could be applied for studying the biological function of PLA2G1B as well as in identifying drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

5

SUMMARY OF THE INVENTION

Accordingly, the inventors herein have discovered 4 novel polymorphic sites in the PLA2G1B gene. These polymorphic sites (PS) correspond to the following nucleotide positions in Figure 1: 3845 (PS1), 6060 (PS3), 6844 (PS4) and 9531 (PS5). The polymorphisms at these sites are guanine or adenine at PS1, adenine or guanine at PS3, guanine or adenine at PS4 and guanine or adenine at PS5. In addition, the inventors have determined the identity of the alleles at these sites, as well as at the previously identified site at nucleotide position 3968 (PS2), in a human reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: African descent, Asian, Caucasian and Hispanic/Latino. From this information, the inventors deduced a set of haplotypes and haplotype pairs for PS1-PS5 in the PLA2G1B gene, which are shown below in Tables 4 and 3, respectively. Each of these PLA2G1B haplotypes defines a naturally-occurring isoform (also referred to herein as an "isogene") of the PLA2G1B gene that exists in the human population. The frequency with which each haplotype and haplotype pair occurs within the total reference population and within each of the four major population groups included in the reference population was also determined.

Thus, in one embodiment, the invention provides a method, composition and kit for genotyping the PLA2G1B gene in an individual. The genotyping method comprises identifying the nucleotide pair that is present at one or more polymorphic sites selected from the group consisting of PS1, PS3, PS4 and PS5 in both copies of the PLA2G1B gene from the individual. A genotyping composition of the invention comprises an oligonucleotide probe or primer which is designed to specifically hybridize to a target region containing, or adjacent to, one of these novel PLA2G1B polymorphic sites. A genotyping kit of the invention comprises a set of oligonucleotides designed to genotype each of these novel PLA2G1B polymorphic sites. In a preferred embodiment, the genotyping kit comprises a set of oligonucleotides designed to genotype each of PS1-PS5. The genotyping method, composition, and kit are useful in determining whether an individual has one of the haplotypes in Table 4 below or has one of the haplotype pairs in Table 3 below.

The invention also provides a method for haplotyping the PLA2G1B gene in an individual. In one embodiment, the haplotyping method comprises determining, for one copy of the PLA2G1B gene, the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS3, PS4 and PS5. In another embodiment, the haplotyping method comprises determining whether one copy of the individual's PLA2G1B gene is defined by one of the PLA2G1B haplotypes shown in Table 4, below, or a sub-haplotype thereof. In a preferred embodiment, the haplotyping method comprises determining whether both copies of the individual's PLA2G1B gene are defined by

one of the PLA2G1B haplotype pairs shown in Table 3 below, or a sub-haplotype pair thereof. The method for establishing the PLA2G1B haplotype or haplotype pair of an individual is useful for improving the efficiency and reliability of several steps in the discovery and development of drugs for treating diseases associated with PLA2G1B activity, e.g., pancreatitis and pancreatic cancer.

5 For example, the haplotyping method can be used by the pharmaceutical research scientist to validate PLA2G1B as a candidate target for treating a specific condition or disease predicted to be associated with PLA2G1B activity. Determining for a particular population the frequency of one or more of the individual PLA2G1B haplotypes or haplotype pairs described herein will facilitate a decision on whether to pursue PLA2G1B as a target for treating the specific disease of interest. In particular, if variable PLA2G1B activity is associated with the disease, then one or more PLA2G1B haplotypes or haplotype pairs will be found at a higher frequency in disease cohorts than in appropriately genetically matched controls. Conversely, if each of the observed PLA2G1B haplotypes are of similar frequencies in the disease and control groups, then it may be inferred that variable PLA2G1B activity has little, if any, involvement with that disease. In either case, the pharmaceutical research scientist can, without *a priori* knowledge as to the phenotypic effect of any PLA2G1B haplotype or haplotype pair, apply the information derived from detecting PLA2G1B haplotypes in an individual to decide whether modulating PLA2G1B activity would be useful in treating the disease.

The claimed invention is also useful in screening for compounds targeting PLA2G1B to treat a specific condition or disease predicted to be associated with PLA2G1B activity. For example, detecting which of the PLA2G1B haplotypes or haplotype pairs disclosed herein are present in individual members of a population with the specific disease of interest enables the pharmaceutical scientist to screen for a compound(s) that displays the highest desired agonist or antagonist activity for each of the most frequent PLA2G1B isoforms present in the disease population. Thus, without requiring any *a priori* knowledge of the phenotypic effect of any particular PLA2G1B haplotype or haplotype pair, the claimed haplotyping method provides the scientist with a tool to identify lead compounds that are more likely to show efficacy in clinical trials.

The method for haplotyping the PLA2G1B gene in an individual is also useful in the design of clinical trials of candidate drugs for treating a specific condition or disease predicted to be associated with PLA2G1B activity. For example, instead of randomly assigning patients with the disease of interest to the treatment or control group as is typically done now, determining which of the PLA2G1B haplotype(s) disclosed herein are present in individual patients enables the pharmaceutical scientist to distribute PLA2G1B haplotypes and/or haplotype pairs evenly to treatment and control groups, thereby reducing the potential for bias in the results that could be introduced by a larger frequency of a PLA2G1B haplotype or haplotype pair that had a previously unknown association with response to the drug being studied in the trial. Thus, by practicing the claimed invention, the scientist can more confidently rely on the information learned from the trial, without first determining the phenotypic effect of any PLA2G1B haplotype or haplotype pair.

In another embodiment, the invention provides a method for identifying an association

between a trait and a PLA2G1B genotype, haplotype, or haplotype pair for one or more of the novel polymorphic sites described herein. The method comprises comparing the frequency of the PLA2G1B genotype, haplotype, or haplotype pair in a population exhibiting the trait with the frequency of the PLA2G1B genotype or haplotype in a reference population. A higher frequency of the PLA2G1B genotype, haplotype, or haplotype pair in the trait population than in the reference population indicates the trait is associated with the PLA2G1B genotype, haplotype, or haplotype pair. In preferred embodiments, the trait is susceptibility to a disease, severity of a disease, the staging of a disease or response to a drug. In a particularly preferred embodiment, the PLA2G1B haplotype is selected from the haplotypes shown in Table 4, or a sub-haplotype thereof. Such methods have applicability in developing diagnostic tests and therapeutic treatments for pancreatitis and pancreatic cancer.

In yet another embodiment, the invention provides an isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the PLA2G1B gene or a fragment thereof. The reference sequence comprises the contiguous sequences shown in Figure 1 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of adenine at PS1, guanine at PS3, adenine at PS4 and adenine at PS5. In a preferred embodiment, the polymorphic variant comprises an additional polymorphism of guanine at PS2.

A particularly preferred polymorphic variant is an isogene of the PLA2G1B gene. A PLA2G1B isogene of the invention comprises guanine or adenine at PS1, adenine or guanine at PS2, adenine or guanine at PS3, guanine or adenine at PS4 and guanine or adenine at PS5. The invention also provides a collection of PLA2G1B isogenes, referred to herein as a PLA2G1B genome anthology.

In another embodiment, the invention provides a polynucleotide comprising a polymorphic variant of a reference sequence for a PLA2G1B cDNA or a fragment thereof. The reference sequence comprises SEQ ID NO:2 (Fig.2) and the polymorphic cDNA comprises at least one polymorphism selected from the group consisting of adenine at a position corresponding to nucleotide 294 and adenine at a position corresponding to nucleotide 365. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a PLA2G1B isogene defined by haplotypes 1c-3c.

Polynucleotides complementary to these PLA2G1B genomic and cDNA variants are also provided by the invention. It is believed that polymorphic variants of the PLA2G1B gene will be useful in studying the expression and function of PLA2G1B, and in expressing PLA2G1B protein for use in screening for candidate drugs to treat diseases related to PLA2G1B activity.

In other embodiments, the invention provides a recombinant expression vector comprising one of the polymorphic genomic variants operably linked to expression regulatory elements as well as a recombinant host cell transformed or transfected with the expression vector. The recombinant vector and host cell may be used to express PLA2G1B for protein structure analysis and drug binding studies.

In yet another embodiment, the invention provides a polypeptide comprising a polymorphic variant of a reference amino acid sequence for the PLA2G1B protein. The reference amino acid sequence comprises SEQ ID NO:3 (Fig.3) and the polymorphic variant comprises histidine at a position corresponding to amino acid position 122. A polymorphic variant of PLA2G1B is useful in

studying the effect of the variation on the biological activity of PLA2G1B as well as on the binding affinity of candidate drugs targeting PLA2G1B for the treatment of pancreatitis and pancreatic cancer.

The present invention also provides antibodies that recognize and bind to the above polymorphic PLA2G1B protein variant. Such antibodies can be utilized in a variety of diagnostic and prognostic formats and therapeutic methods.

The present invention also provides nonhuman transgenic animals comprising one of the PLA2G1B polymorphic genomic variants described herein and methods for producing such animals. The transgenic animals are useful for studying expression of the PLA2G1B isogenes *in vivo*, for *in vivo* screening and testing of drugs targeted against PLA2G1B protein, and for testing the efficacy of therapeutic agents and compounds for pancreatitis and pancreatic cancer in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the PLA2G1B gene. The computer system comprises a computer processing unit; a display; and a database containing the polymorphism data. The polymorphism data includes the polymorphisms, the genotypes and the haplotypes identified for the PLA2G1B gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing PLA2G1B haplotypes organized according to their evolutionary relationships.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a reference sequence for the PLA2G1B gene (Genaissance Reference No. 4762747; contiguous lines), with the start and stop positions of each region of coding sequence indicated with a bracket ([or]) and the numerical position below the sequence and the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence. SEQ ID NO:1 is equivalent to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25). SEQ ID NO:24 is a modified version of SEQ ID NO:1 that shows the context sequence of each polymorphic site, PS1-PS5, in a uniform format to facilitate electronic searching. For each polymorphic site, SEQ ID NO:24 contains a block of 60 bases of the nucleotide sequence encompassing the centrally-located polymorphic site at the 30th position, followed by 60 bases of unspecified sequence to represent that each PS is separated by genomic sequence whose composition is defined elsewhere herein.

Figure 2 illustrates a reference sequence for the PLA2G1B coding sequence (contiguous lines; SEQ ID NO:2), with the polymorphic site(s) and polymorphism(s) identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence.

Figure 3 illustrates a reference sequence for the PLA2G1B protein (contiguous lines; SEQ ID NO:3), with the variant amino acid(s) caused by the polymorphism(s) of Figure 2 positioned below the polymorphic site in the sequence.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is based on the discovery of novel variants of the PLA2G1B gene. As described in more detail below, the inventors herein discovered 6 isogenes of the PLA2G1B gene by characterizing the PLA2G1B gene found in genomic DNAs isolated from an Index Repository that contains immortalized cell lines from one chimpanzee and 93 human individuals. The human individuals included a reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: Caucasian (21 individuals), African descent (20 individuals), Asian (20 individuals), or Hispanic/Latino (18 individuals). To the extent possible, the members of this reference population were organized into population subgroups by their self-identified ethnogeographic origin as shown in Table 1 below.

Table 1. Population Groups in the Index Repository

Population Group	Population Subgroup	No. of Individuals
African descent		20
	Sierra Leone	1
Asian		20
	Burma	1
	China	3
	Japan	6
	Korea	1
	Philippines	5
	Vietnam	4
Caucasian		21
	British Isles	3
	British Isles/Central	4
	British Isles/Eastern	1
	Central/Eastern	1
	Eastern	3
	Central/Mediterranean	1
	Mediterranean	2
	Scandinavian	2
Hispanic/Latino		18
	Caribbean	8
	Caribbean (Spanish Descent)	2
	Central American (Spanish Descent)	1
	Mexican American	4
	South American (Spanish Descent)	3

In addition, the Index Repository contains three unrelated indigenous American Indians (one from each of North, Central and South America), one three-generation Caucasian family (from the CEPH Utah cohort) and one two-generation African-American family.

The PLA2G1B isogenes present in the human reference population are defined by haplotypes for 5 polymorphic sites in the PLA2G1B gene, 4 of which are believed to be novel. The PLA2G1B polymorphic sites identified by the inventors are referred to as PS1-PS5 to designate the order in which they are located in the gene (see Table 2 below), with the novel polymorphic sites referred to as PS1,

PS3, PS4 and PS5. Using the genotypes identified in the Index Repository for PS1-PS5 and the methodology described in the Examples below, the inventors herein also determined the pair of haplotypes for the PLA2G1B gene present in individual human members of this repository. The human genotypes and haplotypes found in the repository for the PLA2G1B gene include those shown in Tables 3 and 4, respectively. The polymorphism and haplotype data disclosed herein are useful for validating whether PLA2G1B is a suitable target for drugs to treat pancreatitis and pancreatic cancer, screening for such drugs and reducing bias in clinical trials of such drugs.

In the context of this disclosure, the following terms shall be defined as follows unless otherwise indicated:

Allele - A particular form of a genetic locus, distinguished from other forms by its particular nucleotide sequence.

Candidate Gene - A gene which is hypothesized to be responsible for a disease, condition, or the response to a treatment, or to be correlated with one of these.

Gene - A segment of DNA that contains all the information for the regulated biosynthesis of an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

Genotype - An unphased 5' to 3' sequence of nucleotide pair(s) found at one or more polymorphic sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype as described below.

Full-genotype - The unphased 5' to 3' sequence of nucleotide pairs found at all polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

Sub-genotype - The unphased 5' to 3' sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a pair of homologous chromosomes in a single individual.

Genotyping - A process for determining a genotype of an individual.

Haplotype - A 5' to 3' sequence of nucleotides found at one or more polymorphic sites in a locus on a single chromosome from a single individual. As used herein, haplotype includes a full-haplotype and/or a sub-haplotype as described below.

Full-haplotype - The 5' to 3' sequence of nucleotides found at all polymorphic sites examined herein in a locus on a single chromosome from a single individual.

Sub-haplotype - The 5' to 3' sequence of nucleotides seen at a subset of the polymorphic sites examined herein in a locus on a single chromosome from a single individual.

Haplotype pair - The two haplotypes found for a locus in a single individual.

Haplotyping - A process for determining one or more haplotypes in an individual and includes use of family pedigrees, molecular techniques and/or statistical inference.

Haplotype data - Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a population; frequency of each haplotype in that or other populations, and any known associations

between one or more haplotypes and a trait.

Isoform – A particular form of a gene, mRNA, cDNA or the protein encoded thereby, distinguished from other forms by its particular sequence and/or structure.

Isogene – One of the isoforms of a gene found in a population. An isogene contains all of the polymorphisms present in the particular isoform of the gene.

Isolated – As applied to a biological molecule such as RNA, DNA, oligonucleotide, or protein, isolated means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally, the term "isolated" is not intended to refer to a complete absence of such material or to absence of water, buffers, or salts, unless they are present in amounts that substantially interfere with the methods of the present invention.

Locus - A location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature.

Naturally-occurring – A term used to designate that the object it is applied to, e.g., naturally-occurring polynucleotide or polypeptide, can be isolated from a source in nature and which has not been intentionally modified by man.

Nucleotide pair – The nucleotides found at a polymorphic site on the two copies of a chromosome from an individual.

Phased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, phased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is known.

Polymorphic site (PS) – A position within a locus at which at least two alternative sequences are found in a population, the most frequent of which has a frequency of no more than 99%.

Polymorphic variant – A gene, mRNA, cDNA, polypeptide or peptide whose nucleotide or amino acid sequence varies from a reference sequence due to the presence of a polymorphism in the gene.

Polymorphism – The sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

Polymorphism data – Information concerning one or more of the following for a specific gene: location of polymorphic sites; sequence variation at those sites; frequency of polymorphisms in one or more populations; the different genotypes and/or haplotypes determined for the gene; frequency of one or more of these genotypes and/or haplotypes in one or more populations; any known association(s) between a trait and a genotype or a haplotype for the gene.

Polymorphism Database – A collection of polymorphism data arranged in a systematic or methodical way and capable of being individually accessed by electronic or other means.

Polynucleotide – A nucleic acid molecule comprised of single-stranded RNA or DNA or comprised of complementary, double-stranded DNA.

Population Group – A group of individuals sharing a common ethnogeographic origin.

Reference Population – A group of subjects or individuals who are predicted to be representative of the genetic variation found in the general population. Typically, the reference population represents the genetic variation in the population at a certainty level of at least 85%, preferably at least 90%, more preferably at least 95% and even more preferably at least 99%.

Single Nucleotide Polymorphism (SNP) – Typically, the specific pair of nucleotides observed at a single polymorphic site. In rare cases, three or four nucleotides may be found.

Subject – A human individual whose genotypes or haplotypes or response to treatment or disease state are to be determined.

Treatment – A stimulus administered internally or externally to a subject.

Unphased – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, unphased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is not known.

As discussed above, information on the identity of genotypes and haplotypes for the PLA2G1B gene of any particular individual as well as the frequency of such genotypes and haplotypes in any particular population of individuals is expected to be useful for a variety of drug discovery and development applications. Thus, the invention also provides compositions and methods for detecting the novel PLA2G1B polymorphisms and haplotypes identified herein.

The compositions comprise at least one PLA2G1B genotyping oligonucleotide. In one embodiment, a PLA2G1B genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that is located close to, or that contains, one of the novel polymorphic sites described herein. As used herein, the term “oligonucleotide” refers to a polynucleotide molecule having less than about 100 nucleotides. A preferred oligonucleotide of the invention is 10 to 35 nucleotides long. More preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan. The oligonucleotide may be comprised of any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives, and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995), pages 617-620). Oligonucleotides of the invention may be prepared by chemical synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the art, including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags and the like.

Genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of a PLA2G1B polynucleotide, i.e., a PLA2G1B isogene. As used herein, specific

hybridization means the oligonucleotide forms an anti-parallel double-stranded structure with the target region under certain hybridizing conditions, while failing to form such a structure when incubated with a non-target region or a non-PLA2G1B polynucleotide under the same hybridizing conditions. Preferably, the oligonucleotide specifically hybridizes to the target region under

5 conventional high stringency conditions. The skilled artisan can readily design and test oligonucleotide probes and primers suitable for detecting polymorphisms in the PLA2G1B gene using the polymorphism information provided herein in conjunction with the known sequence information for the PLA2G1B gene and routine techniques.

A nucleic acid molecule such as an oligonucleotide or polynucleotide is said to be a "perfect"

10 or "complete" complement of another nucleic acid molecule if every nucleotide of one of the molecules is complementary to the nucleotide at the corresponding position of the other molecule. A nucleic acid molecule is "substantially complementary" to another molecule if it hybridizes to that molecule with sufficient stability to remain in a duplex form under conventional low-stringency conditions. Conventional hybridization conditions are described, for example, by Sambrook J. et al.,
15 in *Molecular Cloning, A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989) and by Haymes, B.D. et al. in *Nucleic Acid Hybridization, A Practical Approach*, IRL Press, Washington, D.C. (1985). While perfectly complementary oligonucleotides are preferred for detecting polymorphisms, departures from complete complementarity are contemplated where such departures do not prevent the molecule from specifically hybridizing to the target region. For example,
20 an oligonucleotide primer may have a non-complementary fragment at its 5' end, with the remainder of the primer being complementary to the target region. Alternatively, non-complementary nucleotides may be interspersed into the oligonucleotide probe or primer as long as the resulting probe or primer is still capable of specifically hybridizing to the target region.

Preferred genotyping oligonucleotides of the invention are allele-specific oligonucleotides. As
25 used herein, the term allele-specific oligonucleotide (ASO) means an oligonucleotide that is able, under sufficiently stringent conditions, to hybridize specifically to one allele of a gene, or other locus, at a target region containing a polymorphic site while not hybridizing to the corresponding region in another allele(s). As understood by the skilled artisan, allele-specificity will depend upon a variety of readily optimized stringency conditions, including salt and formamide concentrations, as well as
30 temperatures for both the hybridization and washing steps. Examples of hybridization and washing conditions typically used for ASO probes are found in Kogan et al., "Genetic Prediction of Hemophilia A" in *PCR Protocols, A Guide to Methods and Applications*, Academic Press, 1990 and Ruaño et al., 87 *Proc. Natl. Acad. Sci. USA* 6296-6300, 1990. Typically, an ASO will be perfectly complementary to one allele while containing a single mismatch for another allele.

35 Allele-specific oligonucleotides of the invention include ASO probes and ASO primers. ASO probes which usually provide good discrimination between different alleles are those in which a central position of the oligonucleotide probe aligns with the polymorphic site in the target region (e.g., approximately the 7th or 8th position in a 15mer, the 8th or 9th position in a 16mer, and the 10th or 11th

position in a 20mer). An ASO primer of the invention has a 3' terminal nucleotide, or preferably a 3' penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. ASO probes and primers hybridizing to either the coding or noncoding strand are contemplated by the invention.

ASO probes and primers listed below use the appropriate nucleotide symbol (R= G or A, Y= T or C, M= A or C, K= G or T, S= G or C, and W= A or T; WIPO standard ST.25) at the position of the polymorphic site to represent the two alternative allelic variants observed at that polymorphic site.

A preferred ASO probe for detecting PLA2G1B gene polymorphisms comprises a nucleotide sequence, listed 5' to 3', selected from the group consisting of:

TGCTAAGRACAAATG (SEQ ID NO:4) and its complement,
ATAACATRTTTTCAC (SEQ ID NO:5) and its complement,
CATACTCRTGCTCTG (SEQ ID NO:6) and its complement, and
TGCGACCRCAACGCT (SEQ ID NO:7) and its complement.

A preferred ASO primer for detecting PLA2G1B gene polymorphisms comprises a nucleotide sequence, listed 5' to 3', selected from the group consisting of:

AACAGTTGCTAAGRA (SEQ ID NO:8); GTTATGCATTTGTYC (SEQ ID NO:9);
CTAAGGATAACATRT (SEQ ID NO:10); CAAGAGGTGAAAAYA (SEQ ID NO:11);
CCTATTCATACTCRT (SEQ ID NO:12); CCGAGCCAGAGCAYG (SEQ ID NO:13);
TGCAACTGCGACCRC (SEQ ID NO:14); and GATGGCAGCGTTGYG (SEQ ID NO:15).

Other genotyping oligonucleotides of the invention hybridize to a target region located one to several nucleotides downstream of one of the novel polymorphic sites identified herein. Such oligonucleotides are useful in polymerase-mediated primer extension methods for detecting one of the novel polymorphisms described herein and therefore such genotyping oligonucleotides are referred to herein as "primer-extension oligonucleotides". In a preferred embodiment, the 3'-terminus of a primer-extension oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately adjacent to the polymorphic site.

A particularly preferred oligonucleotide primer for detecting PLA2G1B gene polymorphisms by primer extension terminates in a nucleotide sequence, listed 5' to 3', selected from the group consisting of:

AGTTGCTAAG (SEQ ID NO:16); ATGCATTTGT (SEQ ID NO:17);
AGGATAACAT (SEQ ID NO:18); GAGGTGAAAA (SEQ ID NO:19);
ATTCATACTC (SEQ ID NO:20); AGCCAGAGCA (SEQ ID NO:21);
AACTGCGACC (SEQ ID NO:22); and GGCAGCGTTG (SEQ ID NO:23).

In some embodiments, a composition contains two or more differently labeled genotyping oligonucleotides for simultaneously probing the identity of nucleotides at two or more polymorphic sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a

polymorphic site.

PLA2G1B genotyping oligonucleotides of the invention may also be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, e.g., WO 98/20020 and WO 98/20019). Such immobilized genotyping oligonucleotides may be used in a variety of polymorphism detection assays, including but not limited to probe hybridization and polymerase extension assays. Immobilized PLA2G1B genotyping oligonucleotides of the invention may comprise an ordered array of oligonucleotides designed to rapidly screen a DNA sample for polymorphisms in multiple genes at the same time.

In another embodiment, the invention provides a kit comprising at least two genotyping oligonucleotides packaged in separate containers. The kit may also contain other components such as hybridization buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as PCR.

The above described oligonucleotide compositions and kits are useful in methods for genotyping and/or haplotyping the PLA2G1B gene in an individual. As used herein, the terms "PLA2G1B genotype" and "PLA2G1B haplotype" mean the genotype or haplotype contains the nucleotide pair or nucleotide, respectively, that is present at one or more of the novel polymorphic sites described herein and may optionally also include the nucleotide pair or nucleotide present at one or more additional polymorphic sites in the PLA2G1B gene. The additional polymorphic sites may be currently known polymorphic sites or sites that are subsequently discovered.

One embodiment of the genotyping method involves isolating from the individual a nucleic acid sample comprising the two copies of the PLA2G1B gene, or a fragment thereof, that are present in the individual, and determining the identity of the nucleotide pair at one or more polymorphic sites selected from the group consisting of PS1, PS3, PS4 and PS5 in the two copies to assign a PLA2G1B genotype to the individual. As will be readily understood by the skilled artisan, the two "copies" of a gene in an individual may be the same allele or may be different alleles. In a preferred embodiment of the genotyping method, the identity of the nucleotide pair at PS2 is also determined. In a particularly preferred embodiment, the genotyping method comprises determining the identity of the nucleotide pair at each of PS1-PS5.

Typically, the nucleic acid sample is isolated from a biological sample taken from the individual, such as a blood sample or tissue sample. Suitable tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. The nucleic acid sample may be comprised of genomic DNA, mRNA, or cDNA and, in the latter two cases, the biological sample must be obtained from a tissue in which the PLA2G1B gene is expressed. Furthermore it will be understood by the skilled artisan that mRNA or cDNA preparations would not be used to detect polymorphisms located in introns or in 5' and 3' untranslated regions. If a PLA2G1B gene fragment is isolated, it must contain the polymorphic site(s) to be genotyped.

One embodiment of the haplotyping method comprises isolating from the individual a nucleic acid sample containing only one of the two copies of the PLA2G1B gene, or a fragment thereof, that is present in the individual and determining in that copy the identity of the nucleotide at one or more polymorphic sites selected from the group consisting of PS1, PS3, PS4 and PS5 in that copy to assign a PLA2G1B haplotype to the individual. The nucleic acid may be isolated using any method capable of separating the two copies of the PLA2G1B gene or fragment such as one of the methods described above for preparing PLA2G1B isogenes, with targeted *in vivo* cloning being the preferred approach. As will be readily appreciated by those skilled in the art, any individual clone will only provide haplotype information on one of the two PLA2G1B gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional PLA2G1B clones will need to be examined. Typically, at least five clones should be examined to have more than a 90% probability of haplotyping both copies of the PLA2G1B gene in an individual. In some embodiments, the haplotyping method also comprises identifying the nucleotide at PS2. In a particularly preferred embodiment, the nucleotide at each of PS1-PS5 is identified.

In another embodiment, the haplotyping method comprises determining whether an individual has one or more of the PLA2G1B haplotypes shown in Table 4. This can be accomplished by identifying, for one or both copies of the individual's PLA2G1B gene, the phased sequence of nucleotides present at each of PS1-PS5. The present invention also contemplates that typically only a subset of PS1-PS5 will need to be directly examined to assign to an individual one or more of the haplotypes shown in Table 4. This is because at least one polymorphic site in a gene is frequently in strong linkage disequilibrium with one or more other polymorphic sites in that gene (Drysdale, CM et al. 2000 *PNAS* 97:10483-10488; Rieder MJ et al. 1999 *Nature Genetics* 22:59-62). Two sites are said to be in linkage disequilibrium if the presence of a particular variant at one site enhances the predictability of another variant at the second site (Stephens, JC 1999, *Mol. Diag.* 4:309-317). Techniques for determining whether any two polymorphic sites are in linkage disequilibrium are well-known in the art (Weir B.S. 1996 *Genetic Data Analysis II*, Sinauer Associates, Inc. Publishers, Sunderland, MA).

In a preferred embodiment, a PLA2G1B haplotype pair is determined for an individual by identifying the phased sequence of nucleotides at one or more polymorphic sites selected from the group consisting of PS1, PS3, PS4 and PS5 in each copy of the PLA2G1B gene that is present in the individual. In a particularly preferred embodiment, the haplotyping method comprises identifying the phased sequence of nucleotides at each of PS1-PS5 in each copy of the PLA2G1B gene. When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, it is also envisioned that if the two copies are labeled with different tags, or are otherwise separately distinguishable or identifiable, it could be possible in some cases to perform the method in the same container. For example, if first and second copies of the gene are labeled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labeled with yet a third different fluorescent dye is used to assay the

polymorphic site(s), then detecting a combination of the first and third dyes would identify the polymorphism in the first gene copy while detecting a combination of the second and third dyes would identify the polymorphism in the second gene copy.

In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic site(s) may be determined by amplifying a target region(s) containing the polymorphic site(s) directly from one or both copies of the PLA2G1B gene, or a fragment thereof, and the sequence of the amplified region(s) determined by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic site in individuals who are homozygous at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example, where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for an individual homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

The target region(s) may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR) (U.S. Patent No. 4,965,188), ligase chain reaction (LCR) (Barany et al., *Proc. Natl. Acad. Sci. USA* 88:189-193, 1991; WO90/01069), and oligonucleotide ligation assay (OLA) (Landegren et al., *Science* 241:1077-1080, 1988).

Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Patent No. 5,130,238; EP 329,822; U.S. Patent No. 5,169,766, WO89/06700) and isothermal methods (Walker et al., *Proc. Natl. Acad. Sci. USA* 89:392-396, 1992).

A polymorphism in the target region may also be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically, allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labeled probe pairs, with one member of the pair showing a perfect match to one variant of a target sequence and the other member showing a perfect match to a different variant. In some embodiments, more than one polymorphic site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of the polymorphic sites being detected.

Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Allele-

specific oligonucleotides may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibers, chips, dishes, and beads.

5 The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

The genotype or haplotype for the PLA2G1B gene of an individual may also be determined by hybridization of a nucleic acid sample containing one or both copies of the gene, or fragment(s) thereof, to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would
10 contain a battery of allele-specific oligonucleotides representing each of the polymorphic sites to be included in the genotype or haplotype.

The identity of polymorphisms may also be determined using a mismatch detection technique, including but not limited to the RNase protection method using riboprobes (Winter et al., *Proc. Natl. Acad. Sci. USA* 82:7575, 1985; Meyers et al., *Science* 230:1242, 1985) and proteins which recognize
15 nucleotide mismatches, such as the E. coli mutS protein (Modrich, P. *Ann. Rev. Genet.* 25:229-253, 1991). Alternatively, variant alleles can be identified by single strand conformation polymorphism (SSCP) analysis (Orita et al., *Genomics* 5:874-879, 1989; Humphries et al., in *Molecular Diagnosis of Genetic Diseases*, R. Elles, ed., pp. 321-340, 1996) or denaturing gradient gel electrophoresis (DGGE) (Wartell et al., *Nucl. Acids Res.* 18:2699-2706, 1990; Sheffield et al., *Proc. Natl. Acad. Sci. USA*
20 86:232-236, 1989).

A polymerase-mediated primer extension method may also be used to identify the polymorphism(s). Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Patent 5,679,524. Related methods are disclosed in WO91/02087, WO90/09455,
25 WO95/17676, U.S. Patent Nos. 5,302,509, and 5,945,283. Extended primers containing a polymorphism may be detected by mass spectrometry as described in U.S. Patent No. 5,605,798. Another primer extension method is allele-specific PCR (Ruaño et al., *Nucl. Acids Res.* 17:8392, 1989; Ruaño et al., *Nucl. Acids Res.* 19, 6877-6882, 1991; WO 93/22456; Turki et al., *J. Clin. Invest.* 95:1635-1641, 1995). In addition, multiple polymorphic sites may be investigated by simultaneously
30 amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in Wallace et al. (WO89/10414).

In addition, the identity of the allele(s) present at any of the novel polymorphic sites described herein may be indirectly determined by genotyping another polymorphic site that is in linkage disequilibrium with the polymorphic site that is of interest. Polymorphic sites in linkage
35 disequilibrium with the presently disclosed polymorphic sites may be located in regions of the gene or in other genomic regions not examined herein. Genotyping of a polymorphic site in linkage disequilibrium with the novel polymorphic sites described herein may be performed by, but is not limited to, any of the above-mentioned methods for detecting the identity of the allele at a polymorphic

site.

In another aspect of the invention, an individual's PLA2G1B haplotype pair is predicted from its PLA2G1B genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying a PLA2G1B
 5 genotype for the individual at two or more PLA2G1B polymorphic sites described herein, enumerating all possible haplotype pairs which are consistent with the genotype, accessing data containing PLA2G1B haplotype pairs identified in a reference population, and assigning a haplotype pair to the individual that is consistent with the data. In one embodiment, the reference haplotype pairs include the PLA2G1B haplotype pairs shown in Table 3.

10 Generally, the reference population should be composed of randomly-selected individuals representing the major ethnogeographic groups of the world. A preferred reference population for use in the methods of the present invention comprises an approximately equal number of individuals from Caucasian, African-descent, Asian and Hispanic-Latino population groups with the minimum number of each group being chosen based on how rare a haplotype one wants to be guaranteed to see. For
 15 example, if one wants to have a $q\%$ chance of not missing a haplotype that exists in the population at a $p\%$ frequency of occurring in the reference population, the number of individuals (n) who must be sampled is given by $2n = \log(1-q)/\log(1-p)$ where p and q are expressed as fractions. A preferred reference population allows the detection of any haplotype whose frequency is at least 10% with about 99% certainty and comprises about 20 unrelated individuals from each of the four population groups
 20 named above. A particularly preferred reference population includes a 3-generation family representing one or more of the four population groups to serve as controls for checking quality of haplotyping procedures.

In a preferred embodiment, the haplotype frequency data for each ethnogeographic group is examined to determine whether it is consistent with Hardy-Weinberg equilibrium. Hardy-Weinberg
 25 equilibrium (D.L. Hartl et al., Principles of Population Genomics, Sinauer Associates (Sunderland, MA), 3rd Ed., 1997) postulates that the frequency of finding the haplotype pair H_1 / H_2 is equal to $p_{H-W}(H_1 / H_2) = 2p(H_1)p(H_2)$ if $H_1 \neq H_2$ and $p_{H-W}(H_1 / H_2) = p(H_1)p(H_2)$ if $H_1 = H_2$. A statistically significant difference between the observed and expected haplotype frequencies could be due to one or more factors including significant inbreeding in the population group, strong selective
 30 pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size does not reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping method such as, for
 35 example, CLASPER System[™] technology (U.S. Patent No. 5,866,404), single molecule dilution, or allele-specific long-range PCR (Michalotos-Beloin et al., *Nucleic Acids Res.* 24:4841-4843, 1996).

In one embodiment of this method for predicting a PLA2G1B haplotype pair for an individual,

the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. Alternatively, the haplotype pair in an individual may be predicted from the individual's genotype for that gene using reported methods (e.g., Clark et al. 1990 *Mol Bio Evol* 7:111-22) or through a commercial haplotyping service such as offered by Genaissance Pharmaceuticals, Inc. (New Haven, CT). In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., *supra*). A preferred process for predicting PLA2G1B haplotype pairs from PLA2G1B genotypes is described in U.S. Provisional Application Serial No. 60/198,340 and the corresponding International Application, PCT/US01/12831.

The invention also provides a method for determining the frequency of a PLA2G1B genotype, haplotype, or haplotype pair in a population. The method comprises, for each member of the population, determining the genotype or the haplotype pair for the novel PLA2G1B polymorphic sites described herein, and calculating the frequency any particular genotype, haplotype, or haplotype pair is found in the population. The population may be a reference population, a family population, a same sex population, a population group, or a trait population (e.g., a group of individuals exhibiting a trait of interest such as a medical condition or response to a therapeutic treatment).

In another aspect of the invention, frequency data for PLA2G1B genotypes, haplotypes, and/or haplotype pairs are determined in a reference population and used in a method for identifying an association between a trait and a PLA2G1B genotype, haplotype, or haplotype pair. The trait may be any detectable phenotype, including but not limited to susceptibility to a disease or response to a treatment. The method involves obtaining data on the frequency of the genotype(s), haplotype(s), or haplotype pair(s) of interest in a reference population as well as in a population exhibiting the trait. Frequency data for one or both of the reference and trait populations may be obtained by genotyping or haplotyping each individual in the populations using one of the methods described above. The haplotypes for the trait population may be determined directly or, alternatively, by the predictive genotype to haplotype approach described above. In another embodiment, the frequency data for the reference and/or trait populations is obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is accessible by a computer. Once the frequency data is obtained, the frequencies of the genotype(s), haplotype(s), or haplotype pair(s) of interest in the reference and trait populations are

compared. In a preferred embodiment, the frequencies of all genotypes, haplotypes, and/or haplotype pairs observed in the populations are compared. If a particular PLA2G1B genotype, haplotype, or haplotype pair is more frequent in the trait population than in the reference population at a statistically significant amount, then the trait is predicted to be associated with that PLA2G1B genotype, haplotype or haplotype pair. Preferably, the PLA2G1B genotype, haplotype, or haplotype pair being compared in the trait and reference populations is selected from the full-genotypes and full-haplotypes shown in Tables 3 and 4, or from sub-genotypes and sub-haplotypes derived from these genotypes and haplotypes.

In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting PLA2G1B or response to a therapeutic treatment for a medical condition. As used herein, "medical condition" includes but is not limited to any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment is desirable, and includes previously and newly identified diseases and other disorders. As used herein the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and adverse response (i.e., side effects).

In order to deduce a correlation between clinical response to a treatment and a PLA2G1B genotype, haplotype, or haplotype pair, it is necessary to obtain data on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This clinical data may be obtained by analyzing the results of a clinical trial that has already been run and/or the clinical data may be obtained by designing and carrying out one or more new clinical trials. As used herein, the term "clinical trial" means any research study designed to collect clinical data on responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enroll subjects.

It is preferred that the individuals included in the clinical population have been graded for the existence of the medical condition of interest. This is important in cases where the symptom(s) being presented by the patients can be caused by more than one underlying condition, and where treatment of the underlying conditions are not the same. An example of this would be where patients experience breathing difficulties that are due to either asthma or respiratory infections. If both sets were treated with an asthma medication, there would be a spurious group of apparent non-responders that did not actually have asthma. These people would affect the ability to detect any correlation between haplotype and treatment outcome. This grading of potential patients could employ a standard physical exam or one or more lab tests. Alternatively, grading of patients could use haplotyping for situations where there is a strong correlation between haplotype pair and disease susceptibility or severity.

The therapeutic treatment of interest is administered to each individual in the trial population and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses and that the investigator will choose the number of responder groups (e.g., low, medium, high) made up by the various responses. In addition, the PLA2G1B gene for each individual in the trial population is

genotyped and/or haplotyped, which may be done before or after administering the treatment.

After both the clinical and polymorphism data have been obtained, correlations between individual response and PLA2G1B genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their PLA2G1B genotype or haplotype (or haplotype pair) (also referred to as a polymorphism group), and then the averages and standard deviations of clinical responses exhibited by the members of each polymorphism group are calculated.

These results are then analyzed to determine if any observed variation in clinical response between polymorphism groups is statistically significant. Statistical analysis methods which may be used are described in L.D. Fisher and G. vanBelle, "Biostatistics: A Methodology for the Health Sciences", Wiley-Interscience (New York) 1993. This analysis may also include a regression calculation of which polymorphic sites in the PLA2G1B gene give the most significant contribution to the differences in phenotype. One regression model useful in the invention is described in PCT Application Serial No. PCT/US00/17540, entitled "Methods for Obtaining and Using Haplotype Data".

A second method for finding correlations between PLA2G1B haplotype content and clinical responses uses predictive models based on error-minimizing optimization algorithms. One of many possible optimization algorithms is a genetic algorithm (R. Judson, "Genetic Algorithms and Their Uses in Chemistry" in Reviews in Computational Chemistry, Vol. 10, pp. 1-73, K. B. Lipkowitz and D. B. Boyd, eds. (VCH Publishers, New York, 1997). Simulated annealing (Press et al., "Numerical Recipes in C: The Art of Scientific Computing", Cambridge University Press (Cambridge) 1992, Ch. 10), neural networks (E. Rich and K. Knight, "Artificial Intelligence", 2nd Edition (McGraw-Hill, New York, 1991, Ch. 18), standard gradient descent methods (Press et al., *supra*, Ch. 10), or other global or local optimization approaches (see discussion in Judson, *supra*) could also be used. Preferably, the correlation is found using a genetic algorithm approach as described in PCT Application Serial No. PCT/US00/17540.

Correlations may also be analyzed using analysis of variation (ANOVA) techniques to determine how much of the variation in the clinical data is explained by different subsets of the polymorphic sites in the PLA2G1B gene. As described in PCT Application Serial No.

PCT/US00/17540, ANOVA is used to test hypotheses about whether a response variable is caused by or correlated with one or more traits or variables that can be measured (Fisher and vanBelle, *supra*, Ch. 10).

From the analyses described above, a mathematical model may be readily constructed by the skilled artisan that predicts clinical response as a function of PLA2G1B genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the PLA2G1B gene may be the basis for designing a diagnostic method to determine those individuals who will or will not respond to the treatment, or alternatively, will respond

at a lower level and thus may require more treatment, i.e., a greater dose of a drug. The diagnostic method may take one of several forms: for example, a direct DNA test (i.e., genotyping or haplotyping one or more of the polymorphic sites in the PLA2G1B gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the diagnostic test results and the underlying PLA2G1B genotype or haplotype that is in turn correlated with the clinical response. In a preferred embodiment, this diagnostic method uses the predictive haplotyping method described above.

In another embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the PLA2G1B gene or a fragment of the gene which contains at least one of the novel polymorphic sites described herein. The nucleotide sequence of a variant PLA2G1B gene is identical to the reference genomic sequence for those portions of the gene examined, as described in the Examples below, except that it comprises a different nucleotide at one or more of the novel polymorphic sites PS1, PS3, PS4 and PS5, and may also comprise an additional polymorphism of guanine at PS2. Similarly, the nucleotide sequence of a variant fragment of the PLA2G1B gene is identical to the corresponding portion of the reference sequence except for having a different nucleotide at one or more of the novel polymorphic sites described herein. Thus, the invention specifically does not include polynucleotides comprising a nucleotide sequence identical to the reference sequence of the PLA2G1B gene, which is defined by haplotype 4, (or other reported PLA2G1B sequences) or to portions of the reference sequence (or other reported PLA2G1B sequences), except for genotyping oligonucleotides as described above.

The location of a polymorphism in a variant gene or fragment is identified by aligning its sequence against SEQ ID NO:1. The polymorphism is selected from the group consisting of adenine at PS1, guanine at PS3, adenine at PS4 and adenine at PS5. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the PLA2G1B gene which is defined by any one of haplotypes 1- 3 and 5 - 6 shown in Table 4 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the PLA2G1B gene from a human genomic library. The clone may be sequenced to determine the identity of the nucleotides at the novel polymorphic sites described herein. Any particular variant claimed herein could be prepared from this clone by performing *in vitro* mutagenesis using procedures well-known in the art.

PLA2G1B isogenes may be isolated using any method that allows separation of the two "copies" of the PLA2G1B gene present in an individual, which, as readily understood by the skilled artisan, may be the same allele or different alleles. Separation methods include targeted *in vivo* cloning (TIVC) in yeast as described in WO 98/01573, U.S. Patent No. 5,866,404, and U.S. Patent No. 5,972,614. Another method, which is described in U.S. Patent No. 5,972,614, uses an allele specific oligonucleotide in combination with primer extension and exonuclease degradation to generate hemizygous DNA targets. Yet other methods are single molecule dilution (SMD) as described in Ruaño et al., *Proc. Natl. Acad. Sci.* 87:6296-6300, 1990; and allele specific PCR (Ruaño et al., 1989,

supra; Ruaño et al., 1991, *supra*; Michalatos-Beloin et al., *supra*).

The invention also provides PLA2G1B genome anthologies, which are collections of PLA2G1B isogenes found in a given population. The population may be any group of at least two individuals, including but not limited to a reference population, a population group, a family
5 population, a clinical population, and a same sex population. A PLA2G1B genome anthology may comprise individual PLA2G1B isogenes stored in separate containers such as microtest tubes, separate wells of a microtitre plate and the like. Alternatively, two or more groups of the PLA2G1B isogenes in the anthology may be stored in separate containers. Individual isogenes or groups of isogenes in a genome anthology may be stored in any convenient and stable form, including but not limited to in
10 buffered solutions, as DNA precipitates, freeze-dried preparations and the like. A preferred PLA2G1B genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 4 below.

An isolated polynucleotide containing a polymorphic variant nucleotide sequence of the invention may be operably linked to one or more expression regulatory elements in a recombinant
15 expression vector capable of being propagated and expressing the encoded PLA2G1B protein in a prokaryotic or a eukaryotic host cell. Examples of expression regulatory elements which may be used include, but are not limited to, the lac system, operator and promoter regions of phage lambda, yeast promoters, and promoters derived from vaccinia virus, adenovirus, retroviruses, or SV40. Other regulatory elements include, but are not limited to, appropriate leader sequences, termination codons,
20 polyadenylation signals, and other sequences required for the appropriate transcription and subsequent translation of the nucleic acid sequence in a given host cell. Of course, the correct combinations of expression regulatory elements will depend on the host system used. In addition, it is understood that the expression vector contains any additional elements necessary for its transfer to and subsequent replication in the host cell. Examples of such elements include, but are not limited to, origins of
25 replication and selectable markers. Such expression vectors are commercially available or are readily constructed using methods known to those in the art (e.g., F. Ausubel et al., 1987, in "Current Protocols in Molecular Biology", John Wiley and Sons, New York, New York). Host cells which may be used to express the variant PLA2G1B sequences of the invention include, but are not limited to, eukaryotic and mammalian cells, such as animal, plant, insect and yeast cells, and prokaryotic cells,
30 such as *E. coli*, or algal cells as known in the art. The recombinant expression vector may be introduced into the host cell using any method known to those in the art including, but not limited to, microinjection, electroporation, particle bombardment, transduction, and transfection using DEAE-dextran, lipofection, or calcium phosphate (see e.g., Sambrook et al. (1989) in "Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In a preferred aspect,
35 eukaryotic expression vectors that function in eukaryotic cells, and preferably mammalian cells, are used. Non-limiting examples of such vectors include vaccinia virus vectors, adenovirus vectors, herpes virus vectors, and baculovirus transfer vectors. Preferred eukaryotic cell lines include COS cells, CHO cells, HeLa cells, NIH/3T3 cells, and embryonic stem cells (Thomson, J. A. et al., 1998

Science 282:1145-1147). Particularly preferred host cells are mammalian cells.

As will be readily recognized by the skilled artisan, expression of polymorphic variants of the PLA2G1B gene will produce PLA2G1B mRNAs varying from each other at any polymorphic site retained in the spliced and processed mRNA molecules. These mRNAs can be used for the preparation of a PLA2G1B cDNA comprising a nucleotide sequence which is a polymorphic variant of the PLA2G1B reference coding sequence shown in Figure 2. Thus, the invention also provides PLA2G1B mRNAs and corresponding cDNAs which comprise a nucleotide sequence that is identical to SEQ ID NO:2 (Fig. 2), or its corresponding RNA sequence, except for having one or more polymorphisms selected from the group consisting of adenine at a position corresponding to nucleotide 294 and adenine at a position corresponding to nucleotide 365. A particularly preferred polymorphic cDNA variant comprises the coding sequence of a PLA2G1B isogene defined by haplotypes 1c-3c. Fragments of these variant mRNAs and cDNAs are included in the scope of the invention, provided they contain the novel polymorphisms described herein. The invention specifically excludes polynucleotides identical to previously identified and characterized PLA2G1B cDNAs and fragments thereof. Polynucleotides comprising a variant RNA or DNA sequence may be isolated from a biological sample using well-known molecular biological procedures or may be chemically synthesized.

As used herein, a polymorphic variant of a PLA2G1B gene fragment comprises at least one novel polymorphism identified herein and has a length of at least 10 nucleotides and may range up to the full length of the gene. Preferably, such fragments are between 100 and 3000 nucleotides in length, and more preferably between 200 and 2000 nucleotides in length, and most preferably between 500 and 1000 nucleotides in length.

In describing the PLA2G1B polymorphic sites identified herein, reference is made to the sense strand of the gene for convenience. However, as recognized by the skilled artisan, nucleic acid molecules containing the PLA2G1B gene may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. Thus, reference may be made to the same polymorphic site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target region containing the polymorphic site. Thus, the invention also includes single-stranded polynucleotides which are complementary to the sense strand of the PLA2G1B genomic variants described herein.

Polynucleotides comprising a polymorphic gene variant or fragment may be useful for therapeutic purposes. For example, where a patient could benefit from expression, or increased expression, of a particular PLA2G1B protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the PLA2G1B isogene encoding that isoform or may already have at least one copy of that isogene.

In other situations, it may be desirable to decrease or block expression of a particular PLA2G1B isogene. Expression of a PLA2G1B isogene may be turned off by transforming a targeted

organ, tissue or cell population with an expression vector that expresses high levels of untranslatable mRNA for the isogene. Alternatively, oligonucleotides directed against the regulatory regions (e.g., promoter, introns, enhancers, 3' untranslated region) of the isogene may block transcription.

Oligonucleotides targeting the transcription initiation site, e.g., between positions -10 and +10 from the start site are preferred. Similarly, inhibition of transcription can be achieved using oligonucleotides that base-pair with region(s) of the isogene DNA to form triplex DNA (see e.g., Gee et al. in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y., 1994). Antisense oligonucleotides may also be designed to block translation of PLA2G1B mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of PLA2G1B mRNA transcribed from a particular isogene.

The oligonucleotides may be delivered to a target cell or tissue by expression from a vector introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, the oligonucleotides may be formulated as a pharmaceutical composition for administration to the patient. Oligoribonucleotides and/or oligodeoxynucleotides intended for use as antisense oligonucleotides may be modified to increase stability and half-life. Possible modifications include, but are not limited to phosphorothioate or 2' O-methyl linkages, and the inclusion of nontraditional bases such as inosine and queosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous nucleases.

The invention also provides an isolated polypeptide comprising a polymorphic variant of the reference PLA2G1B amino acid sequence shown in Figure 3. The location of a variant amino acid in a PLA2G1B polypeptide or fragment of the invention is identified by aligning its sequence against SEQ ID NO:3 (Fig. 3). A PLA2G1B protein variant of the invention comprises an amino acid sequence identical to SEQ ID NO:3 except for having histidine at a position corresponding to amino acid position 122. The invention specifically excludes amino acid sequences identical to those previously identified for PLA2G1B, including SEQ ID NO:3, and previously described fragments thereof. PLA2G1B protein variants included within the invention comprise all amino acid sequences based on SEQ ID NO:3 and having histidine at a position corresponding to amino acid position 122.

The invention also includes PLA2G1B peptide variants, which are any fragments of a PLA2G1B protein variant that contain histidine at a position corresponding to amino acid position 122. A PLA2G1B peptide variant is at least 6 amino acids in length and is preferably any number between 6 and 30 amino acids long, more preferably between 10 and 25, and most preferably between 15 and 20 amino acids long. Such PLA2G1B peptide variants may be useful as antigens to generate antibodies specific for one of the above PLA2G1B isoforms. In addition, the PLA2G1B peptide variants may be useful in drug screening assays.

A PLA2G1B variant protein or peptide of the invention may be prepared by chemical synthesis or by expressing one of the variant PLA2G1B genomic and cDNA sequences as described above. Alternatively, the PLA2G1B protein variant may be isolated from a biological sample of an individual having a PLA2G1B isogene which encodes the variant protein. Where the sample contains

two different PLA2G1B isoforms (i.e., the individual has different PLA2G1B isogenes), a particular PLA2G1B isoform of the invention can be isolated by immunoaffinity chromatography using an antibody which specifically binds to that particular PLA2G1B isoform but does not bind to the other PLA2G1B isoform.

5 The expressed or isolated PLA2G1B protein may be detected by methods known in the art, including Coomassie blue staining, silver staining, and Western blot analysis using antibodies specific for the isoform of the PLA2G1B protein as discussed further below. PLA2G1B variant proteins can be purified by standard protein purification procedures known in the art, including differential precipitation, molecular sieve chromatography, ion-exchange chromatography, isoelectric focusing, gel
10 electrophoresis, affinity and immunoaffinity chromatography and the like. (Ausubel et. al., 1987, In Current Protocols in Molecular Biology John Wiley and Sons, New York, New York). In the case of immunoaffinity chromatography, antibodies specific for a particular polymorphic variant may be used.

A polymorphic variant PLA2G1B gene of the invention may also be fused in frame with a heterologous sequence to encode a chimeric PLA2G1B protein. The non-PLA2G1B portion of the
15 chimeric protein may be recognized by a commercially available antibody. In addition, the chimeric protein may also be engineered to contain a cleavage site located between the PLA2G1B and non-PLA2G1B portions so that the PLA2G1B protein may be cleaved and purified away from the non-PLA2G1B portion.

An additional embodiment of the invention relates to using a novel PLA2G1B protein isoform
20 in any of a variety of drug screening assays. Such screening assays may be performed to identify agents that bind specifically to all known PLA2G1B protein isoforms or to only a subset of one or more of these isoforms. The agents may be from chemical compound libraries, peptide libraries and the like. The PLA2G1B protein or peptide variant may be free in solution or affixed to a solid support. In one embodiment, high throughput screening of compounds for binding to a PLA2G1B variant may
25 be accomplished using the method described in PCT application WO84/03565, in which large numbers of test compounds are synthesized on a solid substrate, such as plastic pins or some other surface, contacted with the PLA2G1B protein(s) of interest and then washed. Bound PLA2G1B protein(s) are then detected using methods well-known in the art.

In another embodiment, a novel PLA2G1B protein isoform may be used in assays to measure
30 the binding affinities of one or more candidate drugs targeting the PLA2G1B protein.

In yet another embodiment, when a particular PLA2G1B haplotype or group of PLA2G1B haplotypes encodes a PLA2G1B protein variant with an amino acid sequence distinct from that of PLA2G1B protein isoforms encoded by other PLA2G1B haplotypes, then detection of that particular PLA2G1B haplotype or group of PLA2G1B haplotypes may be accomplished by detecting expression
35 of the encoded PLA2G1B protein variant using any of the methods described herein or otherwise commonly known to the skilled artisan.

In another embodiment, the invention provides antibodies specific for and immunoreactive with one or more of the novel PLA2G1B variant proteins described herein. The antibodies may be

either monoclonal or polyclonal in origin. The PLA2G1B protein or peptide variant used to generate the antibodies may be from natural or recombinant sources or produced by chemical synthesis using synthesis techniques known in the art. If the PLA2G1B protein variant is of insufficient size to be antigenic, it may be conjugated, complexed, or otherwise covalently linked to a carrier molecule to enhance the antigenicity of the peptide. Examples of carrier molecules, include, but are not limited to, albumins (e.g., human, bovine, fish, ovine), and keyhole limpet hemocyanin (Basic and Clinical Immunology, 1991, Eds. D.P. Stites, and A.I. Terr, Appleton and Lange, Norwalk Connecticut, San Mateo, California).

In one embodiment, an antibody specifically immunoreactive with one of the novel protein isoforms described herein is administered to an individual to neutralize activity of the PLA2G1B isoform expressed by that individual. The antibody may be formulated as a pharmaceutical composition which includes a pharmaceutically acceptable carrier.

Antibodies specific for and immunoreactive with one of the novel protein isoforms described herein may be used to immunoprecipitate the PLA2G1B protein variant from solution as well as react with PLA2G1B protein isoforms on Western or immunoblots of polyacrylamide gels on membrane supports or substrates. In another preferred embodiment, the antibodies will detect PLA2G1B protein isoforms in paraffin or frozen tissue sections, or in cells which have been fixed or unfixed and prepared on slides, coverslips, or the like, for use in immunocytochemical, immunohistochemical, and immunofluorescence techniques.

In another embodiment, an antibody specifically immunoreactive with one of the novel PLA2G1B protein variants described herein is used in immunoassays to detect this variant in biological samples. In this method, an antibody of the present invention is contacted with a biological sample and the formation of a complex between the PLA2G1B protein variant and the antibody is detected. As described, suitable immunoassays include radioimmunoassay, Western blot assay, immunofluorescent assay, enzyme linked immunoassay (ELISA), chemiluminescent assay, immunohistochemical assay, immunocytochemical assay, and the like (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Press, New York, New York; Current Protocols in Molecular Biology, 1987, Eds. Ausubel et al., John Wiley and Sons, New York, New York). Standard techniques known in the art for ELISA are described in Methods in Immunodiagnosis, 2nd Ed., Eds. Rose and Bigazzi, John Wiley and Sons, New York 1980; and Campbell et al., 1984, Methods in Immunology, W.A. Benjamin, Inc.). Such assays may be direct, indirect, competitive, or noncompetitive as described in the art (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Pres, NY, NY; and Oellrich, M., 1984, J. Clin. Chem. Clin. Biochem., 22:895-904). Proteins may be isolated from test specimens and biological samples by conventional methods, as described in Current Protocols in Molecular Biology, supra.

Exemplary antibody molecules for use in the detection and therapy methods of the present invention are intact immunoglobulin molecules, substantially intact immunoglobulin molecules, or

those portions of immunoglobulin molecules that contain the antigen binding site. Polyclonal or monoclonal antibodies may be produced by methods conventionally known in the art (e.g., Kohler and Milstein, 1975, *Nature*, 256:495-497; Campbell Monoclonal Antibody Technology, the Production and Characterization of Rodent and Human Hybridomas, 1985, In: *Laboratory Techniques in Biochemistry and Molecular Biology*, Eds. Burdon et al., Volume 13, Elsevier Science Publishers, Amsterdam). The antibodies or antigen binding fragments thereof may also be produced by genetic engineering. The technology for expression of both heavy and light chain genes in *E. coli* is the subject of PCT patent applications, publication number WO 901443, WO 901443 and WO 9014424 and in Huse et al., 1989, *Science*, 246:1275-1281. The antibodies may also be humanized (e.g., Queen, C. et al. 1989 *Proc. Natl. Acad. Sci. USA* 86:10029).

Effect(s) of the polymorphisms identified herein on expression of PLA2G1B may be investigated by preparing recombinant cells and/or nonhuman recombinant organisms, preferably recombinant animals, containing a polymorphic variant of the PLA2G1B gene. As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA into PLA2G1B protein (including codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

To prepare a recombinant cell of the invention, the desired PLA2G1B isogene may be introduced into the cell in a vector such that the isogene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the PLA2G1B isogene is introduced into a cell in such a way that it recombines with the endogenous PLA2G1B gene present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired PLA2G1B gene polymorphism. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner. Examples of cells into which the PLA2G1B isogene may be introduced include, but are not limited to, continuous culture cells, such as COS, NIH/3T3, and primary or culture cells of the relevant tissue type, i.e., they express the PLA2G1B isogene. Such recombinant cells can be used to compare the biological activities of the different protein variants.

Recombinant nonhuman organisms, i.e., transgenic animals, expressing a variant PLA2G1B gene are prepared using standard procedures known in the art. Preferably, a construct comprising the variant gene is introduced into a nonhuman animal or an ancestor of the animal at an embryonic stage, i.e., the one-cell stage, or generally not later than about the eight-cell stage. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill in

the art. One method involves transfecting into the embryo a retrovirus constructed to contain one or more insulator elements, a gene or genes of interest, and other components known to those skilled in the art to provide a complete shuttle vector harboring the insulated gene(s) as a transgene, see e.g., U.S. Patent No. 5,610,053. Another method involves directly injecting a transgene into the embryo. A third method involves the use of embryonic stem cells. Examples of animals into which the PLA2G1B isogenes may be introduced include, but are not limited to, mice, rats, other rodents, and nonhuman primates (see "The Introduction of Foreign Genes into Mice" and the cited references therein, In: Recombinant DNA, Eds. J.D. Watson, M. Gilman, J. Witkowski, and M. Zoller; W.H. Freeman and Company, New York, pages 254-272). Transgenic animals stably expressing a human PLA2G1B isogene and producing human PLA2G1B protein can be used as biological models for studying diseases related to abnormal PLA2G1B expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

An additional embodiment of the invention relates to pharmaceutical compositions for treating disorders affected by expression or function of a novel PLA2G1B isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel PLA2G1B isogenes; an antisense oligonucleotide directed against one of the novel PLA2G1B isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel PLA2G1B isogene described herein. Preferably, the composition contains the active ingredient in a therapeutically effective amount. By therapeutically effective amount is meant that one or more of the symptoms relating to disorders affected by expression or function of a novel PLA2G1B isogene is reduced and/or eliminated. The composition also comprises a pharmaceutically acceptable carrier, examples of which include, but are not limited to, saline, buffered saline, dextrose, and water. Those skilled in the art may employ a formulation most suitable for the active ingredient, whether it is a polynucleotide, oligonucleotide, protein, peptide or small molecule antagonist. The pharmaceutical composition may be administered alone or in combination with at least one other agent, such as a stabilizing compound. Administration of the pharmaceutical composition may be by any number of routes including, but not limited to oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, intradermal, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

For any composition, determination of the therapeutically effective dose of active ingredient and/or the appropriate route of administration is well within the capability of those skilled in the art. For example, the dose can be estimated initially either in cell culture assays or in animal models. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. The exact dosage will be determined by the practitioner, in light of factors

relating to the patient requiring treatment, including but not limited to severity of the disease state, general health, age, weight and gender of the patient, diet, time and frequency of administration, other drugs being taken by the patient, and tolerance/response to the treatment.

Any or all analytical and mathematical operations involved in practicing the methods of the present invention may be implemented by a computer. In addition, the computer may execute a program that generates views (or screens) displayed on a display device and with which the user can interact to view and analyze large amounts of information relating to the PLA2G1B gene and its genomic variation, including chromosome location, gene structure, and gene family, gene expression data, polymorphism data, genetic sequence data, and clinical data population data (e.g., data on ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations). The PLA2G1B polymorphism data described herein may be stored as part of a relational database (e.g., an instance of an Oracle database or a set of ASCII flat files). These polymorphism data may be stored on the computer's hard drive or may, for example, be stored on a CD-ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer via a network.

Preferred embodiments of the invention are described in the following examples. Other embodiments within the scope of the claims herein will be apparent to one skilled in the art from consideration of the specification or practice of the invention as disclosed herein. It is intended that the specification, together with the examples, be considered exemplary only, with the scope and spirit of the invention being indicated by the claims which follow the examples.

EXAMPLES

The Examples herein are meant to exemplify the various aspects of carrying out the invention and are not intended to limit the scope of the invention in any way. The Examples do not include detailed descriptions for conventional methods employed, such as in the performance of genomic DNA isolation, PCR and sequencing procedures. Such methods are well-known to those skilled in the art and are described in numerous publications, for example, Sambrook, Fritsch, and Maniatis, "Molecular Cloning: A Laboratory Manual", 2nd Edition, Cold Spring Harbor Laboratory Press, USA, (1989).

EXAMPLE 1

This example illustrates examination of various regions of the PLA2G1B gene for polymorphic sites.

Amplification of Target Regions

The following target regions of the PLA2G1B gene were amplified using PCR primer pairs. The primers used for each region are represented below by providing the nucleotide positions of their initial and final nucleotides, which correspond to positions in Figure 1.

PCR Primer Pairs

Fragment No.	Forward Primer	Reverse Primer	PCR Product
Fragment 1	3562-3584	complement of 4187-4164	626 nt
5 Fragment 2	3807-3829	complement of 4283-4260	477 nt
Fragment 3	5491-5514	complement of 6144-6121	654 nt
Fragment 4	6488-6511	complement of 7022-6998	535 nt
Fragment 5	9279-9302	complement of 9899-9877	620 nt

10

These primer pairs were used in PCR reactions containing genomic DNA isolated from immortalized cell lines for each member of the Index Repository. The PCR reactions were carried out under the following conditions:

Reaction volume	= 10 μ l
15 10 x Advantage 2 Polymerase reaction buffer (Clontech)	= 1 μ l
100 ng of human genomic DNA	= 1 μ l
10 mM dNTP	= 0.4 μ l
Advantage 2 Polymerase enzyme mix (Clontech)	= 0.2 μ l
Forward Primer (10 μ M)	= 0.4 μ l
20 Reverse Primer (10 μ M)	= 0.4 μ l
Water	= 6.6 μ l

Amplification profile:

97°C - 2 min.	1 cycle
25 97°C - 15 sec.	} 10 cycles
70°C - 45 sec.	
72°C - 45 sec.	
30 97°C - 15 sec.	} 35 cycles
64°C - 45 sec.	
72°C - 45 sec.	

35 Sequencing of PCR Products

The PCR products were purified using a Whatman/Polyfiltronics 100 μ l 384 well unfilter plate essentially according to the manufacturers protocol. The purified DNA was eluted in 50 μ l of distilled water. Sequencing reactions were set up using Applied Biosystems Big Dye Terminator chemistry essentially according to the manufacturers protocol. The purified PCR products were

40 sequenced in both directions using the primer sets described previously or those represented below by the nucleotide positions of their initial and final nucleotides, which correspond to positions in Figure 1. Reaction products were purified by isopropanol precipitation, and run on an Applied Biosystems 3700 DNA Analyzer.

Sequencing Primer Pairs

Fragment No.	Forward Primer	Reverse Primer
Fragment 1	3612-3632	complement of 4160-4141
Fragment 2	3881-3900	complement of 4228-4208
Fragment 3	5648-5667	complement of 6124-6097
Fragment 4	6532-6551	complement of 6994-6973
Fragment 5	9366-9387	complement of 9839-9820

10 Analysis of Sequences for Polymorphic Sites

Sequence information for a minimum of 80 humans was analyzed for the presence of polymorphisms using the Polyphred program (Nickerson et al., *Nucleic Acids Res.* 14:2745-2751, 1997). The presence of a polymorphism was confirmed on both strands. The polymorphisms and their locations in the PLA2G1B gene are listed in Table 2 below.

15

Table 2. Polymorphic Sites Identified in the PLA2G1B Gene

Polymorphic Site Number	PolyId ^a	Nucleotide Position	Reference Allele	Variant Allele	CDS Variant Position	AA Variant
PS1	4762792	3845	G	A		
PS2 ^R	4762790	3968	A	G		
PS3	4762786	6060	A	G		
PS4	4762784	6844	G	A	294	S98S
PS5	4762770	9531	G	A	365	R122H

^aPolyId is a unique identifier assigned to each PS by Genaissance Pharmaceuticals, Inc.

^RPreviously identified in literature

EXAMPLE 2

This example illustrates analysis of the PLA2G1B polymorphisms identified in the Index Repository for human genotypes and haplotypes.

The different genotypes containing these polymorphisms that were observed in the reference population are shown in Table 3 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below.

Table 3, homozygous positions are indicated by one nucleotide and heterozygous positions are indicated by two nucleotides. Missing nucleotides in any given genotype in Table 3 were inferred based on linkage disequilibrium and/or Mendelian inheritance.

Table 3. Genotypes and Haplotype Pairs Observed for PLA2G1B Gene

Genotype Number	Polymorphic Sites					HAP Pair	
	PS1	PS2	PS3	PS4	PS5		
1	G	A	A	G	G	4	4
2	G	A	A	A	G	2	2
3	G/A	A	A	G/A	G	4	1
4	G	A	A	G	G/A	4	3
5	G/A	A	A	A	G	2	1
6	G	A	A/G	G	G	4	5
7	G	A/G	A	G	G	4	6
8	G	A	A	G/A	G	4	2

The haplotype pairs shown in Table 3 were estimated from the unphased genotypes using a computer-implemented extension of Clark's algorithm (Clark, A.G. 1990 *Mol Bio Evol* 7, 111-122) for assigning haplotypes to unrelated individuals in a population sample, as described in U.S. Provisional Application Serial No. 60/198,340 entitled "A Method and System for Determining Haplotypes from a Collection of Polymorphisms" and the corresponding International Application, PCT/US01/12831. In this method, haplotypes are assigned directly from individuals who are homozygous at all sites or heterozygous at no more than one of the variable sites. This list of haplotypes is then used to deconvolute the unphased genotypes in the remaining (multiply heterozygous) individuals. In our analysis, the list of haplotypes was augmented with haplotypes obtained from two families (one three-generation Caucasian family and one two-generation African-American family).

By following this protocol, it was determined that the Index Repository examined herein and, by extension, the general population contains the 6 human PLA2G1B haplotypes shown in Table 4 below.

A PLA2G1B isogene defined by a full-haplotype shown in Table 4 below comprises the regions of the SEQ ID NOS indicated in Table 4, with their corresponding set of polymorphic locations and identities, which are also set forth in Table 4.

Table 4. Haplotypes of the PLA2G1B Gene.

Haplotype	Number ^a						PS	PS	SEQ	Region
	1	2	3	4	5	6	Number ^b	Position ^c	ID NO. ^d	Examined ^e
A	G	G	G	G	G	G	1	3845	1/24	3562-4283
A	A	A	A	A	A	G	2	3968	1/24	3562-4283
A	A	A	A	A	G	A	3	6060	1/24	5491-6144
A	A	G	G	G	G	G	4	6844	1/24	6488-6998
G	G	A	G	G	G	G	5	9531	1/24	9279-9899

^aAlleles for PLA2G1B haplotypes are presented 5' to 3' in each column

^bPS = polymorphic site;

^cPosition of PS within the indicated SEQ ID NO, with the 1st position number referring to the first SEQ ID NO and the 2nd position number referring to the 2nd SEQ ID NO;

^d1st SEQ ID NO refers to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol; 2nd SEQ ID NO is a modified version of the 1st SEQ ID NO that comprises the context sequence of each polymorphic site, PS1-PS5, to facilitate electronic searching of the haplotypes;

^eRegion examined represents the nucleotide positions defining the start and stop positions within the 1st SEQ ID NO of the sequenced region.

SEQ ID NO:1 refers to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol. SEQ ID NO:24 is a modified version of SEQ ID NO:1 that shows the context sequence of each of PS1-PS5 in a uniform format to facilitate electronic searching of the PLA2G1B haplotypes. For each polymorphic site, SEQ ID NO:24 contains a block of 60 bases of the nucleotide sequence encompassing the centrally-located polymorphic site at the 30th position, followed by 60 bases of unspecified sequence to represent that each polymorphic site is

separated by genomic sequence whose composition is defined elsewhere herein.

Table 5 below shows the percent of chromosomes characterized by a given PLA2G1B haplotype for all unrelated individuals in the Index Repository for which haplotype data was obtained. The percent of these unrelated individuals who have a given PLA2G1B haplotype pair is shown in Table 6. In Tables 5 and 6, the "Total" column shows this frequency data for all of these unrelated individuals, while the other columns show the frequency data for these unrelated individuals categorized according to their self-identified ethnogeographic origin. Abbreviations used in Tables 5 and 6 are AF = African Descent, AS = Asian, CA = Caucasian, HL = Hispanic-Latino, and AM = Native American.

Table 5. Frequency of Observed PLA2G1B Haplotypes In Unrelated Individuals

HAP No.	HAP ID	Total	CA	AF	AS	HL	AM
1	4763541	1.22	0.0	5.0	0.0	0.0	0.0
2	4763540	23.17	16.67	55.0	5.0	13.89	33.33
3	4763544	0.61	0.0	2.5	0.0	0.0	0.0
4	4763539	73.78	83.33	37.5	90.0	86.11	66.67
5	4763542	0.61	0.0	0.0	2.5	0.0	0.0
6	4763543	0.61	0.0	0.0	2.5	0.0	0.0

Table 6. Frequency of Observed PLA2G1B Haplotype Pairs In Unrelated Individuals

HAP1	HAP2	Total	CA	AF	AS	HL	AM
4	4	57.32	71.43	10.0	80.0	72.22	33.33
2	2	8.54	4.76	30.0	0.0	0.0	0.0
4	1	1.22	0.0	5.0	0.0	0.0	0.0
4	3	1.22	0.0	5.0	0.0	0.0	0.0
2	1	1.22	0.0	5.0	0.0	0.0	0.0
4	5	1.22	0.0	0.0	5.0	0.0	0.0
4	6	1.22	0.0	0.0	5.0	0.0	0.0
4	2	28.05	23.81	45.0	10.0	27.78	66.67

The size and composition of the Index Repository were chosen to represent the genetic diversity across and within four major population groups comprising the general United States population. For example, as described in Table 1 above, this repository contains approximately equal sample sizes of African-descent, Asian-American, European-American, and Hispanic-Latino population groups. Almost all individuals representing each group had all four grandparents with the same ethnogeographic background. The number of unrelated individuals in the Index Repository provides a sample size that is sufficient to detect SNPs and haplotypes that occur in the general population with high statistical certainty. For instance, a haplotype that occurs with a frequency of 5% in the general population has a probability higher than 99.9% of being observed in a sample of 80 individuals from the general population. Similarly, a haplotype that occurs with a frequency of 10% in a specific population group has a 99% probability of being observed in a sample of 20 individuals from that population group. In addition, the size and composition of the Index Repository means that the

relative frequencies determined therein for the haplotypes and haplotype pairs of the PLA2G1B gene are likely to be similar to the relative frequencies of these PLA2G1B haplotypes and haplotype pairs in the general U.S. population and in the four population groups represented in the Index Repository.

The genetic diversity observed for the three Native Americans is presented because it is of scientific interest, but due to the small sample size it lacks statistical significance.

In view of the above, it will be seen that the several advantages of the invention are achieved and other advantageous results attained.

As various changes could be made in the above methods and compositions without departing from the scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

All references cited in this specification, including patents and patent applications, are hereby incorporated in their entirety by reference. The discussion of references herein is intended merely to summarize the assertions made by their authors and no admission is made that any reference constitutes prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

What is Claimed is:

1. A method for haplotyping the phospholipase A2, group IB (pancreas) (PLA2G1B) gene of an individual, which comprises determining which of the PLA2G1B haplotypes shown in the table immediately below defines one copy of the individual's PLA2G1B gene, wherein each of the PLA2G1B haplotypes comprises a set of polymorphisms whose locations and identities are set forth in the table immediately below:

Haplotype Number ^a						PS	PS
1	2	3	4	5	6	Number ^b	Position ^c
A	G	G	G	G	G	1	3845
A	A	A	A	A	G	2	3968
A	A	A	A	G	A	3	6060
A	A	G	G	G	G	4	6844
G	G	A	G	G	G	5	9531

^aAlleles for haplotypes are presented 5' to 3' in each column

^bPS = polymorphic site;

^cPosition of PS within SEQ ID NO:1.

2. The method of claim 1, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS5 on the one copy of the individual's PLA2G1B gene.
3. A method for haplotyping the phospholipase A2, group IB (pancreas) (PLA2G1B) gene of an individual, which comprises determining which of the PLA2G1B haplotype pairs shown in the table immediately below defines both copies of the individual's PLA2G1B gene, wherein each of the PLA2G1B haplotype pairs consists of first and second haplotypes which comprise first and second sets of polymorphisms whose locations and identities are set forth in the table immediately below:

Haplotype Pair ^a									PS	PS
4/4	2/2	4/1	4/3	2/1	4/5	4/6	4/2		Number ^b	Position ^c
G/G	G/G	G/A	G/G	G/A	G/G	G/G	G/G		1	3845
A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/A		2	3968
A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A		3	6060
G/G	A/A	G/A	G/G	A/A	G/G	G/G	G/A		4	6844
G/G	G/G	G/G	G/A	G/G	G/G	G/G	G/G		5	9531

^aHaplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5' to 3' as 1st polymorphism/2nd polymorphism in each column;

^bPS = polymorphic site;

^cPosition of PS in SEQ ID NO:1.

4. The method of claim 3, wherein the determining step comprises identifying the phased sequence of nucleotides present at each of PS1-PS5 on both copies of the individual's PLA2G1B gene.
5. A method for genotyping the phospholipase A2, group IB (pancreas) (PLA2G1B) gene of an

individual, comprising determining for the two copies of the PLA2G1B gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites (PS) selected from the group consisting of PS1, PS3, PS4 and PS5, wherein the one or more PS have the location and alternative alleles shown in SEQ ID NO:1.

6. The method of claim 5, wherein the determining step comprises:

- (a) isolating from the individual a nucleic acid mixture comprising both copies of the PLA2G1B gene, or a fragment thereof, that are present in the individual;
- (b) amplifying from the nucleic acid mixture a target region containing the selected polymorphic site;
- (c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;
- (d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and
- (e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

7. The method of claim 5, which comprises determining for the two copies of the PLA2G1B gene present in the individual the identity of the nucleotide pair at each of PS1-PS5.

8. A method for haplotyping the phospholipase A2, group IB (pancreas) (PLA2G1B) gene of an individual which comprises determining, for one copy of the PLA2G1B gene present in the individual, the identity of the nucleotide at two or more polymorphic sites (PS) selected from the group consisting of PS1, PS3, PS4 and PS5, wherein the selected PS have the location and alternative alleles shown in SEQ ID NO:1.

9. The method of claim 8, further comprising determining the identity of the nucleotide at PS2, which has the location and alternative alleles shown in SEQ ID NO:1.

10. The method of claim 8, wherein the determining step comprises:

- (a) isolating from the individual a nucleic acid sample containing only one of the two copies of the PLA2G1B gene, or a fragment thereof, that is present in the individual;
- (b) amplifying from the nucleic acid sample a target region containing the selected polymorphic site;
- (c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;
- (d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the selected polymorphic site; and
- (e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.

11. A method for predicting a haplotype pair for the phospholipase A2, group IB (pancreas) (PLA2G1B) gene of an individual comprising:

- (a) identifying a PLA2G1B genotype for the individual, wherein the genotype comprises the nucleotide pair at two or more polymorphic sites (PS) selected from the group consisting of PS1, PS3, PS4 and PS5, wherein the selected PS have the location and alternative alleles shown in SEQ ID NO:1;
- (b) enumerating all possible haplotype pairs which are consistent with the genotype;
- (c) comparing the possible haplotype pairs to the haplotype pair data set forth in the table immediately below; and
- (d) assigning a haplotype pair to the individual that is consistent with the data

Haplotype Pair ^a								PS	PS
4/4	2/2	4/1	4/3	2/1	4/5	4/6	4/2	Number ^b	Position ^c
G/G	G/G	G/A	G/G	G/A	G/G	G/G	G/G	1	3845
A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/A	2	3968
A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	3	6060
G/G	A/A	G/A	G/G	A/A	G/G	G/G	G/A	4	6844
G/G	G/G	G/G	G/A	G/G	G/G	G/G	G/G	5	9531

^aHaplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5' to 3' as 1st polymorphism/2nd polymorphism in each column;

^bPS = polymorphic site;

^cPosition of PS in SEQ ID NO:1.

12. The method of claim 11, wherein the identified genotype of the individual comprises the nucleotide pair at each of PS1-PS5, which have the location and alternative alleles shown in SEQ ID NO:1.
13. A method for identifying an association between a trait and at least one haplotype or haplotype pair of the phospholipase A2, group IB (pancreas) (PLA2G1B) gene which comprises comparing the frequency of the haplotype or haplotype pair in a population exhibiting the trait with the frequency of the haplotype or haplotype pair in a reference population, wherein the haplotype is selected from haplotypes 1-6 shown in the table presented immediately below, wherein each of the haplotypes comprises a set of polymorphisms whose locations and identities are set forth in the table immediately below:

Haplotype Number ^a						PS	PS
1	2	3	4	5	6	Number ^b	Position ^c
A	G	G	G	G	G	1	3845
A	A	A	A	A	G	2	3968
A	A	A	A	G	A	3	6060
A	A	G	G	G	G	4	6844
G	G	A	G	G	G	5	9531

^aAlleles for haplotypes are presented 5' to 3' in each column

^bPS = polymorphic site;

^cPosition of PS in SEQ ID NO:1;

and wherein the haplotype pair is selected from the haplotype pairs shown in the table immediately below, wherein each of the PLA2G1B haplotype pairs consists of first and second haplotypes which comprise first and second sets of polymorphisms whose locations and identities are set forth in the table immediately below:

Haplotype Pair ^a								PS	PS
4/4	2/2	4/1	4/3	2/1	4/5	4/6	4/2	Number ^b	Position ^c
G/G	G/G	G/A	G/G	G/A	G/G	G/G	G/G	1	3845
A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/A	2	3968
A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	3	6060
G/G	A/A	G/A	G/G	A/A	G/G	G/G	G/A	4	6844
G/G	G/G	G/G	G/A	G/G	G/G	G/G	G/G	5	9531

^aHaplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5' to 3' as 1st polymorphism/2nd polymorphism in each column;

^bPS = polymorphic site;

^cPosition of PS in SEQ ID NO:1;

wherein a higher frequency of the haplotype or haplotype pair in the trait population than in the reference population indicates the trait is associated with the haplotype or haplotype pair.

14. The method of claim 13, wherein the trait is a clinical response to a drug targeting PLA2G1B.
15. An isolated genotyping oligonucleotide for detecting a polymorphism in the phospholipase A2, group IB (pancreas) (PLA2G1B) gene at a polymorphic site (PS) selected from the group consisting of PS1, PS3, PS4 and PS5, wherein the selected PS have the location and alternative alleles shown in SEQ ID NO:1.
16. The isolated genotyping oligonucleotide of claim 15, which is an allele-specific oligonucleotide that specifically hybridizes to an allele of the PLA2G1B gene at a region containing the polymorphic site.
17. The allele-specific oligonucleotide of claim 16, which comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-7, the complements of SEQ ID NOS:4-7, and SEQ ID NOS:8-15.
18. The isolated genotyping oligonucleotide of claim 15, which is a primer-extension oligonucleotide.
19. The primer-extension oligonucleotide of claim 18, which comprises a nucleotide sequence selected from the group consisting of SEQ ID NOS:16-23.
20. A kit for genotyping the phospholipase A2, group IB (pancreas) (PLA2G1B) gene of an individual, which comprises a set of oligonucleotides designed to genotype each of polymorphic sites (PS) PS1, PS3, PS4 and PS5, wherein the selected PS have the location and alternative alleles shown in SEQ ID NO:1.
21. The kit of claim 20, which further comprises oligonucleotides designed to genotype PS2, having the location and alternative alleles shown in SEQ ID NO:1.

22. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:

- (a) a first nucleotide sequence which comprises a phospholipase A2, group IB (pancreas) (PLA2G1B) isogene, wherein the PLA2G1B isogene is selected from the group consisting of isogenes 1- 3 and 5 - 6 shown in the table immediately below and wherein each of the isogenes comprises the regions of SEQ ID NO:1 shown in the table immediately below and wherein each of the isogenes 1-3 and 5 - 6 is further defined by the corresponding set of polymorphisms whose locations and identities are set forth in the table immediately below

Isogene Number ^a						PS	PS	SEQ ID	Region
1	2	3	4	5	6	Number ^b	Position ^c	No. ^d	Examined ^e
A	G	G	G	G	G	1	3845	1	3562-4283
A	A	A	A	A	G	2	3968	1	3562-4283
A	A	A	A	G	A	3	6060	1	5491-6144
A	A	G	G	G	G	4	6844	1	6488-7022
G	G	A	G	G	G	5	9531	1	9279-9899

^aAlleles for isogenes are presented 5' to 3' in each column

^bPS = polymorphic site;

^cPosition of PS in SEQ ID NO:1;

^d1st SEQ ID NO:1 refers to Figure 1, with the two alternative allelic variants of each polymorphic site indicated by the appropriate nucleotide symbol;

^eRegion examined represents the nucleotide positions defining the start and stop positions within SEQ ID NO:1 of the sequenced region.

- (b) a second nucleotide sequence which comprises a fragment of the first nucleotide sequence, wherein the fragment comprises one or more polymorphisms selected from the group consisting of adenine at PS1, guanine at PS3, adenine at PS4 and adenine at PS5, wherein the selected polymorphism has the location set forth in the table immediately above; and
- (c) a third nucleotide sequence which is complementary to the first or second nucleotide sequence.
23. The isolated polynucleotide of claim 22, which is a DNA molecule and comprises both the first and third nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.
24. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 22, wherein the organism expresses a PLA2G1B protein encoded by the first nucleotide sequence.
25. The recombinant nonhuman organism of claim 24, which is a transgenic animal.
26. The isolated polynucleotide of claim 22 which consists of the second nucleotide sequence.
27. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:
- (a) a coding sequence for a phospholipase A2, group IB (pancreas) (PLA2G1B) isogene

wherein the coding sequence is defined by a haplotype selected from the group consisting of 1c-3c shown in the table immediately below and wherein the coding sequence comprises SEQ ID NO:2 except at each of the polymorphic sites which have the locations and polymorphisms set forth in the table immediately below:

Coding Sequence	Haplotype Number ^a	PS Number ^b	PS Position ^c
1c-2c	3c		
A	G	4	294
G	A	5	365

^aAlleles for coding sequence haplotypes are presented 5' to 3' in each column; the numerical portion of the coding sequence haplotype number represents the number of the parent full PLA2G1B haplotype;

^bPS = polymorphic site;

^cPosition of PS in SEQ ID NO:2;

and

- (b) a fragment of the coding sequence, wherein the fragment comprises at least one polymorphism selected from the group consisting of adenine at a position corresponding to nucleotide 294 and adenine at a position corresponding to nucleotide 365, wherein said positions in the coding sequence and the fragment refer to SEQ ID NO:2.
28. A recombinant nonhuman organism transformed or transfected with the isolated polynucleotide of claim 27, wherein the organism expresses a phospholipase A2, group IB (pancreas) (PLA2G1B) protein encoded by the polymorphic variant sequence.
 29. The recombinant nonhuman organism of claim 28, which is a transgenic animal.
 30. An isolated polypeptide comprising an amino acid sequence which is a polymorphic variant of a reference sequence for the phospholipase A2, group IB (pancreas) (PLA2G1B) protein or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:3 and the polymorphic variant comprises one or more variant amino acids selected from the group consisting of histidine at a position corresponding to amino acid position 122.
 31. An isolated monoclonal antibody specific for and immunoreactive with the isolated polypeptide of claim 30.
 32. A method for screening for drugs targeting the isolated polypeptide of claim 30 which comprises contacting the PLA2G1B polymorphic variant with a candidate agent and assaying for binding activity.
 33. A computer system for storing and analyzing polymorphism data for the phospholipase A2, group IB (pancreas) gene, comprising:
 - (a) a central processing unit (CPU);
 - (b) a communication interface;
 - (c) a display device;
 - (d) an input device; and

(e) a database containing the polymorphism data;
wherein the polymorphism data comprises the haplotypes set forth in the table immediately below:

10	Haplotype Number ^a						PS	PS
	1	2	3	4	5	6	Number ^b	Position ^c
	A	G	G	G	G	G	1	3845
	A	A	A	A	A	G	2	3968
	A	A	A	A	G	A	3	6060
15	A	A	G	G	G	G	4	6844
	G	G	A	G	G	G	5	9531

^aAlleles for haplotypes are presented 5' to 3' in each column

^bPS = polymorphic site;

^cPosition of PS in SEQ ID NO:1;

20

and the haplotype pairs set forth in the table immediately below:

	Haplotype Pair ^a								PS	PS
	4/4	2/2	4/1	4/3	2/1	4/5	4/6	4/2	Number ^b	Position ^c
25	G/G	G/G	G/A	G/G	G/A	G/G	G/G	G/G	1	3845
	A/A	A/A	A/A	A/A	A/A	A/A	A/G	A/A	2	3968
	A/A	A/A	A/A	A/A	A/A	A/G	A/A	A/A	3	6060
	G/G	A/A	G/A	G/G	A/A	G/G	G/G	G/A	4	6844
	G/G	G/G	G/G	G/A	G/G	G/G	G/G	G/G	5	9531

30

^aHaplotype pairs are represented as 1st haplotype/2nd haplotype; with alleles of each haplotype shown 5' to 3' as 1st polymorphism/2nd polymorphism in each column;

^bPS = polymorphic site;

^cPosition of PS in SEQ ID NO:1.

34. A genome anthology for the phospholipase A2, group IB (pancreas) (PLA2G1B) gene which comprises PLA2G1B isogenes defined by any one of haplotypes 1-6 set forth in the table shown below:

5	Haplotype Number ^a						PS	PS
	1	2	3	4	5	6	Number ^b	Position ^c
	A	G	G	G	G	G	1	3845
	A	A	A	A	A	G	2	3968
	A	A	A	A	G	A	3	6060
10	A	A	G	G	G	G	4	6844
	G	G	A	G	G	G	5	9531

^aAlleles for haplotypes are presented 5' to 3' in each column

^bPS = polymorphic site;

^cPosition of PS in SEQ ID NO:1.

1/8

POLYMORPHISMS IN THE PLA2G1B GENE

CATAAGAGGG	ACTGAACATT	TTGCTGGGCA	AGGCTTCCCT	TTGCTTGGGC	
AGACTCCACT	CATTCTGGGG	CTGCAGAGGC	AGGACCATTC	AGTCAAGCTG	100
ATGTGGGATT	CTGACCTAAC	CAAGTCCCCC	TCCATTAGTC	CTCATAGCCC	
CCACCTCCCA	TGGGGCAGCC	CTGAGACAGG	CTCTGTGACA	ATCCACAGCA	200
GCCCTGTCCA	ACAGAACCTT	CTGTGATCAT	GGAAACATTG	TGTGGCTGCC	
AATCTGGCAG	CCACTCGCCA	CATGTGTCTA	TGAGCCTTGA	AATGTGGCCA	300
TTGTGACTGA	GAAACTGAAC	TTTTAATGGT	ATTTTCATTTT	TATTTTTTATT	
TTTTTTTTTAT	TTATTTTTGAG	GCAGAGTCTC	ACTCTGTCAC	CCGGGCTGGA	400
GTGCAGTGGC	ACTCGGCTCA	CTGCAAGCTC	CGCCTCCCGG	GTTCACGCCA	
TTCTCCTGCC	TCAGCCTCGG	GAGTACCTGG	GACAACAGGC	ACCCGCCACC	500
ACGCCCGGCT	AATTTTTTGT	ATTTTGTAGTA	GAGATGGGGT	TTCACCATGG	
TCTCGATCTC	CTGACCTCAG	GTGATCCACC	CGCTTCGGCC	TCCCGAAGTG	600
CTGGGACTGC	AGGCATGAGC	CACCACGCCC	GGCCCAGAAA	AGAGATGATT	
AAACATAAAG	CAGCCATGTG	ATGAAATGGC	ACTTTGCCTC	TGTGGTCTTC	700
CTCCCCAAA	CCCATAACTG	TAATCTAATT	ATGAGAAAAA	CACAGGACAA	
TTCCAATAGA	GAGCCAGGTG	CAGTGGTTCA	CGCCTGTAAT	CCCAGCACTT	800
TGGGAGGCTG	AGGCGGGCAG	ATCATGAGGT	CAAGAAATCA	AGACCATCCT	
GGCCAACATG	GTGAAACCCC	GTCTCTACTA	AAAATACAAA	AATTAGCTGG	900
ACGCAGTGGT	GTGCACCTGT	AGTCCAGCT	ACTCGGGAGG	CTGAGGCAGG	
AGAATCATTT	GAACCCGGGA	GGCAGAGGTT	GCAGTGAGCT	GAGATCGCGC	1000
CATGCGACTC	CAGCCTGGTG	ACAGAGTGAG	ACTCCGTCTC	AAAAATAAAT	
AAAAATAAAT	AAATAAAAAAT	TAGCTGGGCG	TGGTGGCACG	TGCCTGTAAT	1100
CCCAGCTACT	CAGAGGCTGA	GGCACAAGAA	TCACCTGAAC	CTGGGAGACA	
GAGATTGCAG	TGAGCCGAGA	TTGTGCCACT	GCACTCCAGC	CTGGGCGACA	1200
GAGTGAGACT	ACAACAAACA	CACACACACA	CACCCACACA	CACACACACA	
CAAATTCCAA	GAGAGGGTCA	TCCTGACCAA	TACTCCTCAA	AACTATCAAG	1300
GTTGCTGGGC	ACAGTGGCTC	ACGCCGTGTA	TCCCAATGCT	TTGGGAGGCT	
TAGATGGGAG	GATCACTTGA	GGCCAGGAGT	TCAAGACCAG	CCTGGGCAAC	1400
ATAGGGAGAC	GCCGTGTCTC	CAAAAAATTTT	TTTGAGACAG	AGTCTCGCTG	
TGTCGCCCCAG	GCCGGAGTAC	AGTGGCGTGA	TCTCGGCTCA	CTGCAAACTC	1500
TGCCTCCTGG	GTTACAGCCA	TTCTTCTGCC	TCAGCCTCCC	AAGTTGCTGA	
GATTACAGGC	ACCCGCCACC	ATGCCAGCT	TATTTTTTGT	ATCTTTAGTA	1600
GAGACAAGGT	TTCACTGTGT	TAGCCAGGAT	GGTCTCCATC	ACCTGACCTC	
GTGATCCGCC	TGCCTCGGTC	TCCCCAAGTG	CTGGGATTAC	AGGTATGAGC	1700
CACCGTGCCT	GGCCCCAAAA	ATTTTTTTTAA	ATTAGCCAGG	TGTGGTGACA	
CATGTCTGTA	GTCCCCACTA	ATCGGGAGGC	TAAGGTGGGA	GGATTGCTTG	1800
AGCCCAGGAG	GTTGAGGCTG	CAGTGAACCTA	TGATCGTGTC	ACTGCACATC	
AGTCTGGGAA	ACAGAGCGAC	ACTTTGTCTC	AAAAAAAAAAAA	AACAGATAAA	1900
TAAATTAAAT	AACCAGGCC	TCCTTATCCC	ACAGGGTTGT	TGTAGAGGTG	
ACATAGGAAC	AGAAGAGCAC	CAAGTTAACC	AATTATAAAT	CTATATAGAG	2000
AGAAGCAGAT	CAGAGGCCAG	GCACAGTGGC	TCATGCCTAT	AATCCCAGCA	
TTTGGGGAGG	CTGAGGAGTG	GATCACCTGA	GGTCAAGAGT	TTGAGACCAG	2100
CCTGACCAAC	ATGGTGAAAC	CTTGTCTCTA	CTAAAAATAC	AAAAATTATC	
CAGGCATGCT	GGCAGGCGCC	TGTAATTCCC	AGCTACACGA	GAGGCTGAGG	2200
CAGGAGAATC	GCTTGAACCT	GGGAGGCGGA	GGTTGCAGTA	AGCCGAGATC	
GTGCCATTGC	ACTCCAGCCT	GGGCGACAAG	AGCGAAACTC	TGTCTCAAAA	2300
AAAAAAGAGA	GAGAGAGAGA	GAAGCAGATT	AGCAGTTACC	AGGGGCTGAG	
GGAGTGTGAC	TGCTAATGGG	TACAGGGTTT	CCTTCTGGAG	TGATAAAAAAT	2400
GTTCTGGAAC	CCCATAGAGG	TGATGGTTTG	ACAACACTGT	GAAGGTACTA	
AATGCCCCCG	AATTGTTTAC	TTAAACGTGG	TTAATGTTAT	GTGAATTTCA	2500
GCTAAACAAT	GTTATGTAGA	TATTTGGCCG	GGCGCGGTGG	CTCACGCCTG	
TAATCCAGC	ATTTTGGGAG	GCCGAGGCAG	GTGGATCACG	AGGTCAGGAG	2600

FIGURE 1A

2/8

ATCGAGACCA	TCCTGGCTAA	TGCGGTGAAA	CCCCATCTCT	ACTAAAAATA	
CAAAAAAAAA	AATTAACCGG	GCGTGGTGGT	GGGTGCCTGT	AGTCCCAGCT	2700
ACTTGGGAGG	CTGAGGCAGG	AGAATGGCAT	GAACCTGGGA	GGCAGAGCTT	
GCAGTGAGCC	AAGATCGCGC	CATTGCACTC	CAGCCTAGGC	AACAGAGCAA	2800
GACTCCGTCT	CAAAAAATAT	ATATAAATAA	ATAGATATGT	GATGTGACAG	
GTTTTTTTTT	GAGATGGAGT	TTTGCTCTTG	TTCCCTAGGC	TGGAGTGCAA	2900
TGGCGTGATC	TCAGCTCACC	GCAACCTCCG	CCTCCAGGGT	TCAAGCCATT	
CTCCTGCCTC	GGCCTCCGGA	GTAGCTGGGA	TTACAGGCAT	AAGCCACCAT	3000
GCCTGGCTAA	TTTTGTGTTT	TTAGTAGAGA	CAGGGTTATT	CCATGTTGGT	
CAGGCTGGTC	TCGAACTCTC	CACCTCAGGT	GATCTGCCAG	CCTCAGCCTC	3100
CCAAAGTGCT	GGGATTACAG	GCATGAGCCA	CCGTGCCTGG	CCTCTGATAT	
GACAGTTCTA	ATGCCCTTTA	GTATTCTATA	ATTCAGACTC	AGGCCTTTGG	3200
AATCCAAAGC	CCAGGTTTTT	CTCACAAACC	CACACTGCAG	AGCGGAGTGG	
TGGAIAAAAA	TAAACCTCT	GCCTTGGAAT	CAGACAGATC	TAAACTGGAG	3300
CCCTATTTTG	TCATTTGCCA	ACTGTGTGAC	CTTGGGCAAG	TTACCGCAAC	
TCTCTGAACC	TGTCTCTTTA	TCTGCAAGGT	GCACGACTGA	TGGGACTATT	3400
CAACCAGACC	CAGTGCACAG	ATTCAGGCAC	TTGATAAGAC	ATTGAGGCTG	
CAGGCAGCGA	TCTTTTTTCT	TTCTTTCTTT	TTTTTTTTTT	TTTTTTGAAA	3500
TAGGGTCTCA	CTCTGCTGCA	GAGGCTCAAT	CACTGTTTCT	TGCAGCCTTG	
ACCTCCCTGG	CTCAAGAGAT	CCTCCCATCT	CAGCCTCCTG	AGTTGCTGGG	3600
ATCACAGGTG	CAATCCACCA	CCACACCTGG	TTAACATTTT	TTTTTTTAGA	
GATGAGGTCT	CTCTATGTTG	CCCAGGCTGC	ACTTCCTTCT	TGTCCTCCCT	3700
ATCCCAGCGT	CCGACTGAAC	TGACGGCTTT	GCTTTCCCCA	ACCAGCCCCG	
GAAGCTGGGC	TGAGTACAAA	GTGGTGGGTA	TGAGGGTCAA	GATTGTAAAG	3800
TCTGAAAAC	CCAGAAACCA	TCCCTTTGGT	TAACAGTTGC	TAAGGACAAA	
			A		
TGCATAACAT	ATTTTCCAGT	GATCCCATGC	TGGCAAATCG	TCAGGGTCAT	3900
TCCTGCAACA	GACAGATTCA	AGGCCAGCCC	CAAACCTCAG	CAAGAGCAAA	
GCAAACTCTC	CAGCCTTATC	TGGGCAGGGT	TGTGTGGAGA	CTGACTATAA	4000
	G				
GACTATACCT	GAGACTGGTC	ATCTCAGTTT	CTTTTCTCAC	CTTGACTGCA	
AGATGAAACT	CCTTGTGCTA	GCTGTGCTGC	TCACAGGTAG	GCAAGTCTCC	4100
[exon 1: 4053..					
..4086]					
CCGGCTCCAC	CCGCCTTTCT	CTCCCAAGTG	AGCTAAGATC	TCACTCCTCT	
GGAATGGGGG	CCACAGGCCA	CAGCAAACAG	GGATGGCCAG	CCCCGCAGTC	4200
TCAATTCGAG	GTTCCCAAGT	GGGCTTAAGG	GCTCCTCTAT	TGGGGTTCCC	
TCAAGGCTGG	CACTTTTTTC	ACCTGCAAGT	CTGAACTCAG	ATTGCCTGAG	4300
CTAAGAAAGC	TTGCCCTTAT	TTTCTTTTTT	CCAGACAGGG	TCTTGCTCTA	
TCACCCAGGC	TGGAGTTCAG	TGGCATGATC	ATAGCTCACC	ACAGCTTCCA	4400
ACTCGTGGGC	TCAAGTGATC	CTCCCACCTT	ACTCAACTAA	GTAGTTAGGC	
CAATCTCCCA	TTTATTTTAT	TTTATTTTAA	TTTTTATTTT	TATTTTACTT	4500
TATTTTATTT	TTGAGACGGG	GCTCACTCTG	TGCCCCAGGC	TGGAGTGCGG	
TGGCGTGATC	TCAGATCACT	ACAACCTCCA	TCTCCTGGGT	TCAAATAATT	4600
CTCTTGCCCT	AGCCTCTCAA	GTAGCTGGGA	CTTGCTAGCTC	TCAAGTAGCT	
GGCACACACC	ACCATGCCCA	GCTAATTTTT	TGTGTGTTTT	TTTTGGTAGA	4700
GACAGGTTTT	CACCATGTTG	GCCAGGCTGG	GTGACCTCCC	TTTTAGATTC	
TCCTCATCTC	GCTCTATTCT	TCCCCTTTCT	AATGCAGTAT	CCAGTTTCCT	4800
TACTTATCAC	ATTATTATT	ATTCTTATTA	TTATTGAGAC	AGAGTCTTGC	
TTGTTCGCCA	AGGCTGGAGT	ACAGTGGTGC	GATCTCGGCT	CACTGCAAGC	4900
TCCACCTGCT	GGGTTCACGC	CATTCTCCCG	CCTCAGCCTC	CCCAGTAGCT	
GGGACTAAAG	GCGCCTGCCA	CCACGCCCCG	CTAATTTTTT	TGTATTTTTA	5000
ATAAAGACGG	GGTTTCATCG	TGTTAGCCAG	GATGGTCTCG	ATCTCATGAC	
CTTGTGATCC	GCCTGCCTCG	GCCTCCCAAA	GTGCTGGGAT	TACAGGCATG	5100

FIGURE 1B

3/8

AGCCACCGTG	CCCGGCCTTA	TCACATTTAT	TATTTATTGT	TTTTCTCTCC	
CACTAGGTTG	TAAGCTCCAT	GAGGTTAGAG	ATTATTATTA	TTATTATTAT	5200
TATTATTATT	ATTATTATTA	TTATTATATC	TGTTCACTGC	TGTATCTCTA	
GCTCCTAGGA	CAGAGCCTGG	CACATAGTAA	GTGCTCAATA	AATATTCACT	5300
GGATAAACAG	TGCAGATAGT	TTAAAACTAT	CTGACCTAGG	GAGGCTGAGG	
CAGGAGAATG	GCGTGAACCC	GGGAAGCAGA	GTTTGCAGTG	AGCTGAAATC	5400
GTGTCACCTG	ACTCCAACCT	GGGCAACAGA	GCAAGACTCC	ATCTCAAAAA	
AAAAAAAAAA	ACTATCAGGC	CTAGCTGGGT	GGCACATGCC	TGTAATCCTA	5500
GCTGAGGCGG	TAGGGTCCCA	GAAGAAGAAG	AAGAAGAAAA	AGAAGAAGAT	
ATATATATAT	ATACACACAC	ACAAAAGATAT	AAACTTTATA	TATATAAAGT	5600
TTTCATTAAA	AAAAAAAAAA	AACCTCTACC	CACTTTCAC	TTACCAGGTT	
CCTGGGTCCA	ACGGTCTTCA	GAGGAGGCAG	CTGGCAGGGG	TCAGGGAGGC	5700
AGCGTGGGAC	CCGAGGGAGC	AGGAAGGCAG	TGTGTCCCCG	GGGTGCTGGC	
AGACCGATTT	GAACCTCTGGC	TATGTCTTCT	TGCAGTGGCC	GCCGCCGACA	5800
[exon 2: 5786..					
GCGGCATCAG	CCCTCGGGCC	GTGTGGCAGT	TCCGCAAAAT	GATCAAGTGC	
GTGATCCCGG	GGAGTGACCC	CTTCTTGGAA	TACAACAAC	ACGGCTGCTA	5900
CTGTGGCTTG	GGGGGCTCAG	GCACCCCGCT	GGATGAAC	GACAAGTAAG	
..5945]					
TGATCCGCCT	GCAGGAAAAT	TGGAGTGCCT	GCCGGGGGCG	GGGTGGGGCA	6000
CACGCCAAGG	ATCTCACGAG	GCATACAAAG	GGGACTTGCA	TATCTGCTAA	
GGATAACATA	TTTTCACCTC	TTGTCAAATA	AACAAATATG	TTCCAAGAGG	6100
G					
ACCCTGTAGC	GAACGCACCC	CGTTAGAGAT	GGAAACAATG	ACCGACGTGC	
AAAACAGTGG	GCGATGCTGC	CCTCCAGTGG	CAGAAATGTAG	CAACAGTAAA	6200
CATCACAGCA	ACTATCCACG	TGTCATTTTC	TAGCAGTGGT	TGTCAC	
CCTTCTGAAT	ACAGGATTTT	ACTGTATTCT	TGCAACCATG	TTAAAAATCG	6300
CTTTCAGGCC	AGGCGCGGTG	GCTCATGCCT	GTAATCCCAG	CACTTTGGGA	
GGCCGAGGCG	GGCGGATCAC	TTGAGGTCAG	GAGTTCGAGA	CCAGCCTGGC	6400
CAACATGGTG	AAACCCTGTC	TCTACTAAAA	AATACAAAAA	TTAGCCGGAC	
ATGGTGGCGA	GCGCCTGTAA	CCCCAGCTAC	TTGGGAGACT	GAGTTGGAGG	6500
TTTCAGTGAG	CCAAGGTCGT	GTCAC	TCCAGCCTGG	GTAACAGAGC	
AACTCTGTCT	CAAAAAAAAA	AAATGCTTTC	AATAAATATA	TGATAAAAGG	6600
ACTTATATTT	TTTCAAGCCA	TAGGATCATT	TCTCCTGAAG	CATCTTGGCG	
AAGTCATCCC	CACCTGTTCC	TGAGAGTGGG	CAGGTGAGGG	CTGACCTATT	6700
GCTCTGCAC	TACTCTATC	TCAGCTGTCC	CTCCCACTTT	CCAGGTGCTG	
[exon 3: 6745..					
CCAGACACAT	GACAACTGCT	ATGACCAGGC	CAAGAAGCTG	GACAGCTGTA	6800
AATTTCTGCT	GGACAACCCG	TACACCCACA	CCTATT	CTCGTGCTCT	
A					
GGCTCGGCAA	TCACCTGTAG	CAGTAGGTTT	ATCCCTTCCT	TGACCTATGA	6900
..6873]					
ATTCTAGTTG	GTTCTCAGTA	GGCCGGGGGG	AAATAATAGT	AACAACAGCC	
ATGATTTAGT	GTTAATTTTC	TTGGTTCTGG	GCAGTGCTC	CTTTAATCCT	7000
CAGAACACA	CTATGGGATA	GGTACAATTA	TCCTCACTTA	ACAGATAAGA	
AAACTGAGGC	TCAGAAGGCT	GAGCTATTTG	CCCAAGATCA	CACAGCTTGT	7100
AAGTGGTGAC	AGTTTGGGTT	TTTTTTTGT	GTTGTTTAGA	GACAGGGTCT	
TGCTCTGTCA	CCCAGGCATG	AGCACAGTGG	TGCAACCATA	GGTCACTGCA	7200
GCCTCAACCT	CCTGAGCTCA	AGGGATCTGC	TGACCTCAGC	CTCCCAAGTA	
GCTGGGACCTA	CGAGCGTGCA	CCACCACGCC	TGGCTAATTA	AAAAAATTTT	7300
TTTGTAGAGA	CTGGGTCTTA	CTACGTTGGC	CAGGCTTGTC	TTAAACTCCT	
GGCTTCAAGC	AATCCTCCTA	CCTTGGCATC	CCAAAGTGCT	GGGATTACAG	7400
GGGTGAGCCA	CCATGTGCGG	CTACTTATTT	CTTTACATTC	CATCTTTCCA	
ATAGAATGTA	AGATCCACAG	AACAGGGATT	ACTGCCTATT	TTCTTCCTTT	7500

FIGURE 1C

4/8

CTTTTTTGGAG	ACAGAGTCTC	ACTTCATCAC	CTCAACCTCC	GTTTCAGCTCA	
CTGCAACCTC	TGCCTCCCGG	GTTCAAGCGA	TTCTCCTGCC	TAAGCCTCCT	7600
GAGTAGCTGG	AATTACAAGC	GTGCACCACC	ATGCTTGGCT	AATTTTTTGT	
ATTTTTAGCA	GAGATGGGGT	TTTACCATGT	TGCCCAGGCT	GGTCTCAAAC	7700
TCCTGACCTC	AAGTGATCTG	CCTGCCTCAG	TCTCCCAAAG	TGCTGGAATT	
ATAGGCGTGA	GTCACGTGTC	CTGGCCGATT	ACTGTCTATT	TTCTTTATTG	7800
CTATATCCCC	AGATCTAGAG	CAGTGTCTGA	CATATAGTAG	GTGCTCAATA	
AATAATTGAT	GAATGCACAG	CCTAGATATA	AACTTTCTTT	TTCTTTTTTT	7900
AAAACAATCT	TGACAACTTT	GCAGAATAAA	TACAATCTTG	CAITCTGCTT	
TTTCACTTAT	CACCTTGTTA	TGACTTTTTTC	ATATTGCCTC	AAACCTTTAT	8000
TGTTACTGTT	TTTTCATTTG	TACTATTTTA	GTCACCTGAAT	AATATGGCTT	
AATTTGCTTA	TACATCCTCC	TGCTCCACTT	TAGAAGGCCA	AATTTACAAA	8100
TCTGATGAAA	GCTATGAACC	CTCTCCCCAG	AGAAATACAC	ACACACACAC	
ACACTCACAC	ACAGTTTTTT	TTTAATGTTT	GCAACTAAGA	CAAGAAACCT	8200
GCATTAGAGG	ATGTTTGTTT	ATATTAATTA	AAAATAACTC	AGTTGGGCAC	
AGTGACTCAA	GCCTGTAACC	ACAGTACTTT	GGAAGTCCAA	GGTGGGTGGA	8300
TCACCTTGAGG	TGAGAAAGTT	GAGACCAGCC	TGGTCAATAT	GGTGAAACCC	
TATCTCTACT	AAAAATACAA	AAATTAGCTG	GGTGTAGTGA	TGCATGCCTG	8400
TAGTCCCAGC	TACTCGGGAG	GCTGAGGCAA	GAGAATTGCT	TGAACCTGGG	
AGGCAGAGGT	TGCAGTGAGC	CGAGATCCCA	CCACTGCAC	CCAGCCTGGG	8500
CGACACAGCG	AGACTCTATC	TCAAAAAAT	AAATAAATAA	AATAAAGGAT	
CGGAGAGAAA	CAAAACTAAT	AAGATTCCTG	AAGGTAAGCA	GAGATACGTA	8600
AATTATATGT	AATAAAGTTT	AAATGCATTT	TAAGTGAAT	CTTATTGTTT	
ATTTTGTTA	TAAAAGTAAA	CAAGCCAAAA	GTAATGCAAC	TTCAAACCTC	8700
ACATAAATAT	CTATTATGGA	AAGTGGAAGG	CATCTATAAT	CCTACTACCC	
AAAGATAACC	AGTTACATAT	TCCTCCAGAT	TTTTGGGGCA	TACACTAGCT	8800
TTTTTTATTT	GGGAAAATTT	CCATGTGCAG	GCATACCTAA	TTTTTCTAAA	
TGTCTATGTA	GTAATCCATT	TAAGGATGTT	CCATAATTTT	TAAAATACAT	8900
GCTTTAAAGT	AGAGAAACTA	GGTTGGGCAT	GGTGGCTCAC	GCCTGTATCC	
CAGCACTTTG	GGAGGCCGAG	GCAAATGGAT	CACTTGAGGT	CCGGAGTTTG	9000
AGACCAGCCT	GGACAACATG	ATGAAACACC	CTCTCTAATA	AAAATACAAA	
AATTAGCTGG	GCATGGTGGC	GAGCACCTGT	AGTCCCAGCT	ACTCAGGAGT	9100
CTGAGGCAGG	AGTATCACTT	GAACCTCAGG	GGCAGAAGTT	GCAGTGAGCT	
GAGATCACGC	CACTGCACTC	CAGCCTAGGC	GACAAAAGGG	AAACTCCGTC	9200
TTAAAAACAA	AACAAAACAA	AAAAACACAG	GATGCCCAGA	TAAATATGAC	
TTTCATATAA	GCAATGGATA	ATTTTTTGGG	GGGTATATGT	CCCAATATT	9300
GCATTGATTG	TTTATCTGAA	AGTCAAATTT	AACTGGGCAT	CCTGATGTAC	
TTGTATTAC	TTAATCTGTC	AGCCCTAAAT	GTGCATCAGT	GGAATGGCTG	9400
CCAGCTTATT	CCAGTTAATT	CTTCTTGCCC	CAGATTGTAC	AAAACAGGGT	
CCACCTTGGC	TCAGTCCTCT	CCTTTCATCC	CTCTCCAGGC	AAAAACAAAG	9500
[exon 4: 9490..					
AGTGTGAGGC	CTTCATTTGC	AACTGCGACC	GCAACGCTGC	CATCTGCTTT	
A					
TCAAAAGCTC	CATATAACAA	GGCACACAAG	AACCTGGACA	CCAAGAAGTA	9600
TTGTCAGAGT	TGAATATCAC	CTCTCAAAAG	CATCACCTCT	ATCTGCCTCA	
..9613]					
TCTCACACTG	TACTCTCCAA	TAAAGCACCT	TGTTGAAAGA	CCTCATGTTT	9700
GGATATTGTT	TTATTCTCTG	TCTATAAACT	AGGTCTCTGC	CTACTCTTTT	
ATTTTATGTT	ATTTATTTTT	TCTAGGTGGA	GTCTTGCTCT	GTGGCCAGG	9800
CTGGAGTGCA	GTGATGCCAC	CTTGCCCTAC	TGCAACCTCC	GCCTCCCGGG	
CTCAAGCAAT	CCTCCCGCCT	CATCCTCCCG	AGTAGCTGGG	ACCATAGGCA	9900
TGCACCACCA	TGCCTGGCTA	ATTTTGTGAT	TTTTTGTAGA	GACAGAGTTT	
CGCCATGTTG	CCCTGGCTGG	TCTCAAACCTC	CTCAGCTTAA	GTGATCTGCC	10000
TGGCTCGGCC	TCGCAAAGTG	TTGGGATTAC	ATGCATGAGC	CGCCGCGCCT	

FIGURE 1D

5/8

GGCTACTCTG	CCTAGTCTTT	TGTGAGTATC	ATTTCTTCCA	GCCTTGGAAG	10100
CTAAGTTGAA	TTAGAAAGAC	ACTTCCAGGA	AGCAAGCAAG	CACCTTGAAA	
CCTGAGTAAT	GATTAACGAT	CACCATCTAC	TGATTATTTA	CTCTGTACCA	10200
GGACTGTGTG	TCCATAAATC	CTCTTGACAG	CCCTGTGAGG	TATTGGCGCT	
ATTAGCAAAT	CTTATTTTCC	TAAGCTGAGG	CTCAATAGGA	GAGGTCACCT	10300
TTCCAATGCT	ATCATCTAGT	AAGCAGCAGA	GAAGGAATTT	GAACCTCGGCA	
AGTCTAACAA	CAGAAAACAC	ATGCTGAACC	ACTGCCCTTC	CCTGCCTGAA	10400
GTGGTAGGCT	TTAGTTTGAG	CCAGACCTTG	CCCCCGTCTC	ATGATTCTGC	
CTCCATTTTC	AACGTGTATTA	AACCATTTTT	CTCAATGAC	TTTCTTTTTT	10500
TTTTTTTTTT	TTGAGATGGA	GTCTCGCTCT	GTCCGCCAGG	CTGGAGTGCA	
GTGCTGCAAT	CTCGGCTCAC	TGCAAGCTCT	GCCTCCCAGG	TTCACGCCAT	10600
TCTCCTGCCT	CAGCTTCCCG	AGTAGCTGGG	TTTACAGGCT	CCTGCCACCA	
CGCCCAGCTA	ATTTTTTGTA	TTTTCAGTAG	AGACGGGGTT	TCACCGTGTT	10700
AGCCAGGATG	GTCTCGATCT	CCTGACCTCG	TGATCCGCCC	GCCTCGGCCT	
CCCAAAGTGC	TGGGATTACA	GGCGTGAGCC	ACCGCACACG	GCCACGACTT	10800
TCTTTTCTAA	ATAAAAGACT	TCACCACACT	CTACAGGCTA	ATTTTGACAC	
TGTAGTCATG	AAATATAATA	AACATTAACA	AGCCGAGCAT	GGCGGCACGC	10900
GCCTATGATC	GTAGCTACTC	AAGAGGCTGA	GGCAGGAGGA	TCTCTTGATC	
CCGGGAGTTT	GAGGCTGCAG	TGAGCTATGA	TCACACCACT	GCACTCCAGC	11000
CTGGGTGAAA	GAGTGAGACC	CTGTTTCAAG	CTACTAGGGA	GGCTGAAAGT	
GAAGGATCCC	CTGAGCCCAG	GAGTTGGAGG	CTGCAGTGAG	CTGTGATCAC	11100
GCCACTGCAC	TCCAGCCTGA	GTGACAGAGA	GAGACACTAT	CTCAAACAAA	
CACACACACA	AAACACAAAC	AAAACAAAAC	AAAACAAAAC	AAAACAAAAC	11200
AAAAAACCAA	TAACAGCTTG	CATTTCTGGA	GCACCTACTG	CATACTTCCT	
TGTTTCGGAGT	TTTCCACATC	TCATCTCATT	AAATGTTCAA	ACCAGCTCTG	11300
TGATATTGAT	ATTTTTGCTC	CCATTTCAIG	GATGTGGAAC	TAAAAATTCA	
GAGAAAGTTAA	GTCATTTGTC	CAAGATCACA	CAAATGGCAA	AATCAGGATT	11400
TGGCCAGGTC	TGTCTGGTGG	CAGTGCCCAA	GCTTTTAAAC	ACTAAGTCAC	
TTCAGCCCAA	TTCTCTATG	AGTATTTATG	ACTACATTTA	CATTGAAATT	11500
CACCAGAACT	AAGCCAGGGA	CAGTGGCTCA	CGCCTGTAAT	CCCAGGACTT	
TGAGAAGTCT	AGGTGGGCAG	ATCACTTGAG	GCCAGGAGTT	TGAGACCAGC	11600
CTGGCCAACA	TGGCAAAACC	CTGTCTCTAC	TAAAAAATAC	AAAAATTAGC	
CGAGTATGGT	GGCATAGGCC	TGTAATCCCA	ACTACTCAGG	AGGCTGACGC	11700
AGGAGAATGG	CTTAAACCAG	GGAGGCTGAG	GTTGTAGTGA	GCTGAGATCA	
CACCACTGCA	CTCCAGCGTG	GGTGACAGAG	TGATACTCGG	CCAAAGAAAG	11800
AAAAAAAAAA	AAGAAATTCA	CCAGAGCAAG	TGGGGAGAGG	CAGAGAAGGG	
AACTGAAGCA	GAAAAGCAAA	GTTGAGGCCA	GGCATGGTGG	CTCACGCCTG	11900
TAATCCCAGT	ACTTTGGGAA	GCCGAAGAGG	GCAGATCACC	TCAGGTCAGG	
AGTTCGAGAC	CAGCCTGGCC	AACATGGCAA	AACCTCATCT	CTACTAAAAA	12000
ATACAAAATT	TAGCTGGGCA	TGGTAGCGTG	CACCTGTAAT	CCCAGCTACT	
CAGGAGGCTG	AGACAGAAGA	ATCACTTGAA	CCTGGGAGGA	AGAGGTTGCA	12100
GTGAGCCAAG	ATAACACCAC	TGCACTCCAG	CCTGGGTGAC	ACAGTGAGAC	
TCTGTCCACT	ATCAAAAGGC	ACCATTAAGC	AGTTTCAGAC	CACTGGAAAC	12200
TGACAGAACA	GCCACCATTT	CTGGGTATGT	GGTGGGCAAT	AGAAGACCTT	
CCATTTGAAG	AAGGGAGTGT	GTGTGGACTT	ATCTTTTGGC	ATAATTTCTC	12300
CCTCATTTAT	TTATTTATTT	ATAATTTTTT	TTTTGAGACA	GATTCTCACT	
CTGTCACCAG	GCTGGAGTGT	AGTGGCGCAA	TCTCGGCTCA	CTGTAACCTC	12400
TGCCTCCCGG	GTTCAAGAGA	TTCCCCTGCC	TCAGCCTCCT	GAGTAGCTGG	
GACTATAGGC	ACGCACCACC	ACGCCAGCT	AATTTTTTGT	ATTTTAGAGA	12500
CAGGGTTTCA	CTATGTTGGC	CAGGATGTTT	TCGATCTCCT	GACCTCGTGA	
TCTGCCCCGAC	TTGGCCTCCC	AAAGTGCTGG	GATTACAGGC	GTGAGCCACC	12600
ACGCCCGGTG	CCTCATTTTA	TTCATCCATT	CAATGATTAT	TTATTGAACA	
CTTACTCTGG	GCCAGGCACC	ACACAAAATG	CTGAGGATAC	AATGTATTCT	12700
GGGCTGGCCA	CAAGACAATG	ATTTAGTGTT	AGTATCTGTC	TACAATTCAG	

FIGURE 1E

6/8

TCTGCTCTGA	GACTTTCTTC	CCTACTCTTC	AAACTGGCAT	CCTTGTATAT	12800
ATTATGCTGT	ATGCTCCAGA	CGGTGTCATG	TCTTTTGCAC	AATTATCTTG	
TCATCCTTGA	AGTTCTATAG	TTGTAGGTAC	TGTTATATCC	ACATCTGATG	12900
GATGAGGAAA	CTGGCTTAAA	GATGCAATGG	ATTGCCAAAA	GTCACTAATT	
CCTGGGTGGC	ATAGCTGACC	TTGGAACCCA	AATCCAGATA	ATTCAGAGAA	13000
GTGCTATCTT	TTTAATTCTT	TTTAATTAAA	ATTAATTCTT	TTTAAGTCTT	
TTTAATTCTT	TGGCAGGCAA	GAATGGGACT	ATTCTCAAAT	TCAAATTCAA	13100
CAAGCATAAA	CCGAAGAACC	CTTCTTTGCA	TATTGTGCTA	TGAAGAATTT	
CAAGGGATGC	AAAATCAAGT	TAGACCTGGC	CTTAATCTTC	AACGAACTGT	13200
GAAGGCCATA	AAAAATATTT	CTTCAGTGGA	AAGAAATTGC	AAGTCCCTGT	
TGGATGGAAG	AAAAGTGGGC	TTTTAGGAAA	ATGGTGCTTA	CTTCTTCTTA	13300
TTTATTTATT	TATTTATTTA	GATGACAGAA	CAAACGAGAG	AGAGAGAGCA	
GTCACAGGGC	TACTAAAGGC	TCTAAAGATA	TCCTCTGATG	TAAAAATCTC	13400
TCTTGGGCCG	GGCGTGGTAG	CTCAGGCCTG	TAAATCCCAG	CACTTTGAGA	
AGCTGAGGCT	TGCTTGAGTC	CAGCCTGGAC	AACACGGTGA	AATCCCGTCT	13500
CTACAAAAAA	TCAGCAAGGT	GTTGCATGGC	TGTAGTCCCA	GCTACTCGGG	
AGGCTGAGAT	GGGAGTATCA	CCTGAGCCCA	GAAGGTTGAG	GCTGCAGTGA	13600
GACTGCACTC	TG				13612

FIGURE 1F

7/8

POLYMORPHISMS IN THE CODING SEQUENCE OF PLA2G1B

ATGAAACTCC	TTGTGCTAGC	TGTGCTGCTC	ACAGTGGCCG	CCGCCGACAG	
CGGCATCAGC	CCTCGGGCCG	TGTGGCAGTT	CCGCAAAATG	ATCAAGTGCG	100
TGATCCCGGG	GAGTGACCCC	TTCTTGGAAT	ACAACAACATA	CGGCTGCTAC	
TGTGGCTTGG	GGGGCTCAGG	CACCCCCGTG	GATGAACTGG	ACAAGTGCTG	200
CCAGACACAT	GACAACTGCT	ATGACCAGGC	CAAGAAGCTG	GACAGCTGTA	
AATTTCTGCT	GGACAACCCG	TACACCCACA	CCTATTCATA	CTCGTGCTCT	300
			A		
GGCTCGGCAA	TCACCTGTAG	CAGCAAAAAC	AAAGAGTGTG	AGGCCTTCAT	
TTGCAACTGC	GACCGCAACG	CTGCCATCTG	CTTTTCAAAA	GCTCCATATA	400
	A				
ACAAGGCACA	CAAGAACCTG	GACACCAAGA	AGTATTGTCA	GAGTTGA	447

FIGURE 2

8/8

ISOFORMS OF THE PLA2G1B PROTEIN

MKLLVLAVLL	TVAAADSGIS	PRAVWQFRKM	IKCVIPGSDP	FLEYNNYGCY	
CGLGGSGTPV	DELDKCCQTH	DNCYDQAKKL	DSCKFLLDNP	YHTYSSYSCS	100
GSAITCSSKN	KECEAFICNC	DRNAAICFSK	APYNKAHKNL	DTKKYCQS	148
	H				

FIGURE 3

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
14 February 2002 (14.02.2002)

PCT

(10) International Publication Number
WO 02/012562 A3

(51) International Patent Classification⁷: **C12Q 1/68**,
C12N 5/10, G06F 17/00, C12N 15/12, 9/20, C07K 16/40

(21) International Application Number: PCT/US01/24663

(22) International Filing Date: 6 August 2001 (06.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US): **GENAIS-
SANCE PHARMACEUTICALS, INC.** [US/US]; A Cor-
poration of the State of Delaware, Five Science Park, New
Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **KAZEMI, Amir**
[US/US]; 53 Southwind Drive, Wallingford, CT US (US).
KLIEM, Stefanie E. [DE/DE]; Kiefernweg 37, 61440
Oberusel (DE). **KOSHY, Beena** [IN/US]; 1298 Hartford
Turnpike, Apt. 11B, North Haven, CT 06473 (US).

(74) Agents: **FIELD, Gisela M.** et al.; Genaissance Pharma-
ceuticals, Inc., Five Science Park, New Haven, CT 06511
(US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,

CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,
ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,
TG).

Published:

- with international search report
- with sequence listing part of description published sepa-
rately in electronic form and available upon request from
the International Bureau

(88) Date of publication of the international search report:
24 July 2003

*For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*

(54) Title: HAPLOTYPES OF THE PLA2G1B GENE

(57) Abstract: Novel genetic variants of the Phospholipase A2, Group IB (Pancreas) (PLA2G1B) gene are described. Various genotypes, haplotypes, and haplotype pairs that exist in the general United States population are disclosed for the PLA2G1B gene. Compositions and methods for haplotyping and/or genotyping the PLA2G1B gene in an individual are also disclosed. Polynucleotides defined by the haplotypes disclosed herein are also described.



WO 02/012562 A3

INTERNATIONAL SEARCH REPORT

Inte. nal Application No
PCT/US 01/24663

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12Q1/68 C12N5/10 G06F17/00 C12N15/12 C12N9/20
C07K16/40

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, BIOSIS, MEDLINE, CHEM ABS Data, EMBASE, EMBL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL [Online] Database accession no. M22970 XP002224243	1-26
Y	abstract	34

X	DATABASE EMBL [Online] Database accession no. AC073930 XP002225091	1-26
Y	abstract	34

	-/--	



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

10 January 2003

Date of mailing of the international search report

30.04.03

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Madlener, M

INTERNATIONAL SEARCH REPORT

Inter. al Application No
PCT/US 01/24663

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SEILHAMER J J ET AL: "PANCREATIC PHOSPHOLIPASE A-2 ISOLATION OF THE HUMAN GENE AND COMPLEMENTARY DNA SPECIES FROM PORCINE PANCREAS AND HUMAN LUNG" DNA (NEW YORK), vol. 5, no. 6, 1986, pages 519-528, XP009003085 ISSN: 0198-0238	1-26
Y	abstract	34
X	--- HALUSHKA M K ET AL: "PATTERNS OF SINGLE-NUCLEOTIDE POLYMORPHISMS IN CANDIDATE GENES FOR BLOOD-PRESSURE HOMEOSTASIS" NATURE GENETICS, NATURE AMERICA, NEW YORK, US, vol. 22, July 1999 (1999-07), pages 239-247, XP000985696 ISSN: 1061-4036	1-26
Y	table 1	34
Y	--- US 5 972 614 A (BENTLEY KEVIN L ET AL) 26 October 1999 (1999-10-26) column 5, line 65 -column 8, line 5	34
A	--- US 6 004 744 A (GOELET PHILIP ET AL) 21 December 1999 (1999-12-21) the whole document	1-26,34
A	--- BEUTEN JOKE ET AL: "Mutation analysis of the human pancreatic phospholipase A-2 gene in a family with distal hereditary motor neuropathy type II linked to 12q24." NEUROSCIENCE LETTERS, vol. 223, no. 1, 1997, pages 69-71, XP002224242 ISSN: 0304-3940 the whole document	1-26,34
A	--- FROSSARD P M ET AL: "TaqI dimorphic site in a human pancreatic phospholipase A2 (PLA2) gene localized on chromosome 12." HUMAN GENETICS. GERMANY JUN 1994, vol. 93, no. 6, June 1994 (1994-06), page 720 XP009002510 ISSN: 0340-6717 the whole document	1-26,34
	--- -/--	

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/24663

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ULBRECHT M ET AL: "Assoziation of beta 2-adrenoreceptor variants with bronchial hyperresponsiveness" AMERICAN JOURNAL OF RESPIRATORY AND CRITICAL CARE MEDICINE, AMERICAN LUNG ASSOCIATION, NEW YORK, NY, US, vol. 161, no. 2, February 2000 (2000-02), pages 469-474, XP002188594 ISSN: 1073-449X cited in the application page 469, right-hand column, paragraph 2 -page 470	1-26,34
A	--- CLAIBORNE STEPHENS J: "SINGLE-NUCLEOTIDE POLYMORPHISMS, HAPLOTYPES AND THEIR RELEVANCE TO PHARMACOGENETICS" MOLECULAR DIAGNOSIS, NAPERVILLE, IL, US, vol. 4, no. 4, 1999, pages 309-317, XP002937520 ISSN: 1084-8592 the whole document	1-26,34
A	--- CLARK A G ET AL: "HAPLOTYPE STRUCTURE AND POPULATION GENETIC INFERENCES FROM NUCLEOTIDE-SEQUENCE VARIATION IN HUMAN LIPOPROTEIN LIPASE" AMERICAN JOURNAL OF HUMAN GENETICS, UNIVERSITY OF CHICAGO PRESS, CHICAGO,, US, vol. 63, 1998, pages 595-612, XP002938535 ISSN: 0002-9297 cited in the application the whole document -----	1-26,34

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 01/24663

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.: 33
because they relate to subject matter not required to be searched by this Authority, namely:
Rule 39.1(vi) PCT - Program for computers
2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-26, 34 (all partially)

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-26, 34 (all partially)

An isolated PLA2G1B polynucleotide comprising a fragment of SEQ.ID.NO.1, wherein the fragment comprises an adenine at PS1; a recombinant nonhuman organism transformed or transfected with said polynucleotide; methods, primers and kits for haplotyping and genotyping the PLA2G1B locus comprising the detection of an adenine or a guanine at PS1; a method for the identification of an association between a trait and a PLA2G1B sequence comprising adenine at PS1; a genome anthology comprising such a PLA2G1B sequence.

2. Claims: 1-26, 34 (all partially)

An isolated PLA2G1B polynucleotide comprising a fragment of SEQ ID No.1, wherein the fragment comprises a guanine at PS3; a recombinant nonhuman organism transformed or transfected with said polynucleotide; methods, primers and kits for haplotyping and genotyping the PLA2G1B locus comprising the detection of an adenine or a guanine at PS3; a method for the identification of an association between a trait and a PLA2G1B sequence comprising adenine at PS3; a genome anthology comprising such a PLA2G1B sequence.

3. Claims: 1-29, 34 (all partially)

An isolated PLA2G1B polynucleotide comprising a fragment of SEQ ID No. 1, wherein the fragment comprises an adenine at PS4; a recombinant nonhuman organism transformed or transfected with said polynucleotide; methods, primers and kits for haplotyping and genotyping the PLA2G1B locus comprising the detection of an adenine or a guanine at PS4; a method for the identification of an association between a trait and a PLA2G1B sequence comprising adenine at PS4; a genome anthology comprising such a PLA2G1B sequence;

an isolated PLA2G1B polynucleotide comprising a fragment of SEQ.ID.NO.2, wherein the fragment comprises an adenine at PS4; a recombinant nonhuman organism transformed or transfected with said polynucleotide.

4. Claims: 1-29, 34 (all partially), 30-32 (entirely)

An isolated PLA2G1B polynucleotide comprising a fragment of SEQ ID No. 1, wherein the fragment comprises an adenine at PS5; a recombinant nonhuman organism transformed or transfected with said polynucleotide; methods, primers and kits for haplotyping and genotyping the PLA2G1B locus comprising the detection of an adenine or a guanine at PS5;

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

a method for the identification of an association between a trait and a PLA2G1B sequence comprising adenine at PS5; a genome anthology comprising such a PLA2G1B sequence;

an isolated PLA2G1B polynucleotide comprising a fragment of SEQ.ID.NO.2, wherein the fragment comprises an adenine at PS5; a recombinant nonhuman organism transformed or transfected with said polynucleotide; an isolated polypeptide comprising a fragment of SEQ.ID.NO.3 comprising a histidine at a position corresponding to amino acid position in SEQ.ID.NO.3; an isolated monoclonal antibody specific for said polypeptide; a method for screening for drugs targeting said polypeptide.

5. Claims: 1-4, 13, 14, 22-29, 34 (all partially)

An isolated PLA2G1B polynucleotide comprising a nucleotide sequence comprising SEQ.ID.NO.1 or SEQ.ID.NO.2 comprising

- a guanine at PS1
- a guanine at PS2
- an adenine at PS3
- a guanine at PS4
- a guanine at PS5;

a recombinant non-human organism transformed or transfected with said polynucleotide; methods for haplotyping the PLA2G1B comprising the determination of such a sequence; method for the identification of an association between a trait and such a PLA2G1B sequence; methods for using such a PLA2G1B sequence for assessing drug trials and treatment regimes, and for screening drugs; a genome anthology comprising such PLA2G1B sequences.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 01/24663

Patent document cited in search report		Publication date		Patent family member(s)		Publication date
US 5972614	A	26-10-1999	US	5866404 A		02-02-1999
			US	6221588 B1		24-04-2001
			US	2002132348 A1		19-09-2002

US 6004744	A	21-12-1999	US	5888819 A		30-03-1999
			AU	660173 B2		15-06-1995
			AU	1584892 A		06-10-1992
			CA	2105060 A1		06-09-1992
			EP	0576558 A1		05-01-1994
			FI	933870 A		03-09-1993
			FI	20010223 A		07-02-2001
			JP	6505394 T		23-06-1994
			NO	933136 A		02-09-1993
			WO	9215712 A1		17-09-1992
			US	2003044779 A1		06-03-2003
			US	5762876 A		09-06-1998
			US	2003044778 A1		06-03-2003
